The Trichotomous Agent

Body, Mind, and Soul in Economic Theory*

Charl Jooste

July 15, 2026

Abstract

Standard economic models represent the agent as body (consumption, health) and mind (human capital, education). The soul does not appear. This paper introduces the *trichotomous agent*, an economic agent with three state variables: body (B), mind (M), and soul (S). The soul achieves a several things: the soul state reshapes time preferences, consumption behavior, social orientation, and moral constraints simultaneously through four channels; patience, satiation, generosity, and honesty. Grace, modeled as an exogenous stochastic process, is the mechanism for spiritual transformation that the agent cannot self-generate. The model nests standard homo economicus as the special case where $S = 0$. We derive seven propositions connecting the soul state to corruption, savings, voluntary redistribution, total factor productivity, forgiveness equilibria, structural unemployment, and the existence of multiple steady states. Calibrated simulations show that a society moving from a low-soul to a high-soul state experiences GDP gains on the order of 60 to 75 percent, a two-thirds reduction in corruption, and a shift from revenge to forgiveness as the dominant social strategy. We argue that Christianity is uniquely positioned among world religions to supply the exogenous grace that the model requires. The model does not reject mainstream economics, but extends it by allowing the soul to be non-zero.

Keywords: spiritual capital, economic growth, utility transformation, grace, social capital, religion and economics, multiple equilibria

JEL Classification: O43, Z12, D91, O11

Author contribution and use of AI. The conception of the trichotomous agent (body, mind, and soul) and all substantive arguments, theological commitments, and final judgments are the author's own. In preparing this paper the author used Claude (Anthropic) as a writing and analysis aid: drafting and editing prose, helping formalize the model and its derivations, generating the simulation code and figures, and surfacing related literature. The author has reviewed and verified the work and takes full responsibility for its entire content, including any AI-assisted portions; the AI is a tool, not an author. I thank colleagues and seminar participants, especially Frederico Gil Sander and Ewa Korczyc, who read drafts and challenged several parts. The views expressed are those of the author alone and do not represent the views of the World Bank Group, its Executive Directors, or the governments they represent.

Contents

1 Introduction **4**

1.1 What this paper does . . . 5
1.2 What this paper does not do . . . 5
1.3 Plan of the paper . . . 5

2 Related Literature **6**

2.1 Economics of religion . . . 6
2.2 Social capital . . . 7
2.3 Institutions and growth . . . 7
2.4 Culture and economic outcomes . . . 8
2.5 Behavioral and preference foundations . . . 9
2.6 Summary: Positioning this paper . . . 11

3 Why Christianity Is Distinct **11**

3.1 Grace is a gift, not an achievement . . . 12
3.2 The ground of grace: corrupted nature, satisfied justice, imputed righteousness . . . 13
3.3 Why not synergistic grace? . . . 18
3.4 Transformation, not compliance: the model's mechanism . . . 20
3.5 Moral dispositions are derivative, not foundational . . . 20
3.6 Relationship to Weber . . . 21
3.7 The dark side: When Christianity fails . . . 22
3.8 The corrosion of works-based religion . . . 23

4 The Model **24**

4.1 Preferences . . . 25
4.2 Budget constraint . . . 26
4.3 The honesty constraint . . . 27
4.4 Laws of motion . . . 27
4.5 Grace . . . 29
4.6 Optimality conditions . . . 31

5 Four Channels and Seven Propositions **33**

5.1 Channel 1: Patience . . . 33
5.2 Channel 2: Satiation . . . 34
5.3 Channel 3: Generosity . . . 34
5.4 Channel 4: Honesty . . . 36
5.5 Additional propositions . . . 36

6 Two Cities, Two Equilibria **39**

6.1 The critical mass and the role of the Church . . . 40
6.2 Historical evidence: Rwanda and South Africa . . . 40

7 Numerical Simulations **40**
7.1 Comparative statics of the soul state . . . 41
7.2 The two-equilibria phase diagram . . . 41
7.3 The welfare landscape . . . 41
7.4 GDP decomposition by soul channel . . . 41
7.5 The forgiveness game . . . 44
7.6 Transition dynamics . . . 44
7.7 Summary of simulation results . . . 45

8 Nesting Result: Homo Economicus as a Special Case **47**

9 A Challenge to the Secular Economist **47**
9.1 "Norms can explain this without God" . . . 48
9.2 "The model is not falsifiable" . . . 49
9.3 "You have a reverse causality problem" . . . 53
9.4 "You are smuggling theology into economics" . . . 54
9.5 "This is a god-of-the-gaps argument" . . . 54

10 Conclusion **56**

A Derivations **60**
A.1 The Lagrangian . . . 60
A.2 First-order conditions . . . 61
A.3 Co-state equations . . . 62
A.4 Steady-state conditions . . . 63
A.5 The consumption Euler equation . . . 63

B Extended Proofs **64**
B.1 Proof of Proposition 1: The threshold policy for corruption . . . 64
B.2 Proof of Proposition 3: g/c is increasing in S . . . 64
B.3 Proof of Proposition 5: The two thresholds of the forgiveness game . . . 64
B.4 Proof of Proposition 7: Multiple equilibria . . . 65

1 Introduction

Why do some nations flourish while others, blessed with the same resources, the same institutions on paper, the same access to global knowledge, decay? The standard answers such as capital accumulation, human capital, technology, institutions explain much but leave a residual. We call it total factor productivity.

A large and successful literature has *endogenized* much of it, decomposing measured TFP into distinct pieces: genuine technological progress and the diffusion of ideas; the efficiency with which resources are allocated across firms and sectors (misallocation); variation in the *utilization* of labor and capital over the business cycle; and plain measurement error in the inputs.

Our claim concerns one component of the residual: neither the cyclical part (the soul is a slow-moving state variable and explains nothing about quarter-to-quarter swings in measured productivity, which are about utilization and demand) nor the measurement-error part. It concerns the *structural, low-frequency* part: why two economies with the same capital, the same years of schooling, and the same access to technology nonetheless differ, persistently and across generations, in how productively those inputs are combined. The endogenization literature lists the proximate mechanisms: misallocation, weak contract enforcement, thin innovation. Our claim is that these mechanisms are not themselves bedrock. Misallocation is what happens when resources flow to whoever can appropriate them rather than whoever can use them best; contracts fail to bind when the people signing them are willing to break them; innovation is thin when patient, honest effort is not rewarded. The soul does not replace the endogenization literature. Instead, it describes the disposition of the people who run the firms, sign the contracts, and staff the bureaus, while *shifting* the mechanisms that literature has identified.

We argue that this part of the residual is attributable to the soul.

We introduce the *trichotomous agent*: an economic agent composed of body, mind, and soul. The body is physical health and vigor. The mind is human capital, e.g., knowledge, skills, cognitive capacity. The soul is spiritual capital that captures moral orientation, character, and relationship with God. The critical departure from standard economics is that the soul does not enter the model as another good to consume or another stock to accumulate. It *transforms the utility function itself.* A change in the soul state changes what the agent wants, how patient the agent is, how generous the agent is, and whether the agent can cheat. It is a meta-variable.

The intellectual ancestry of this idea is ancient. Augustine of Hippo, writing in the early fifth century as the Roman Empire crumbled around him, distinguished between two cities: the *civitas Dei* (City of God) and the *civitas terrena* (City of Man). The City of God is composed of those who love God above self. The City of Man is composed of those who love self above God. These two cities are intermixed in history, coexisting in every society, but they produce radically different outcomes such as different virtues, different institutions, and different destinies (Augustine, 426).

Augustine's two cities map onto our model's two equilibria. *civitas Dei* is the high-soul equilibrium where patience, generosity, honesty, and satiation characterize the representative agent. The low-soul equilibrium is where impatience, greed, corruption, and insatiable desire dominate - it is the economic expression of the *civitas terrena*. The critical insight, which Augustine understood which our model attempts to formalize, is that the transition between

these two states cannot be accomplished by human effort alone. It requires grace.

1.1 What this paper does

We make five contributions:

1. We build a formal dynamic optimization model with three state variables (body, mind, soul) where the soul transforms preferences and the budget set. We derive the full Lagrangian, first-order conditions, and co-state equations.

2. We derive seven propositions connecting the soul state to observable economic outcomes: corruption (Proposition 1), savings and investment (Proposition 2), voluntary redistribution (Proposition 3), total factor productivity (Proposition 4), inter-group forgiveness (Proposition 5), structural unemployment (Proposition 6), and the existence of two stable equilibria (Proposition 7). Each proposition states the assumptions under which it holds, including the boundary cases where the comparative static is weak rather than strict.

3. We show that the model *nests* standard homo economicus as the special case $S = 0$. The trichotomous agent extends mainstream economics.

4. We present calibrated simulations showing that the quantitative effects of the soul state are large: a society moving from the low-soul equilibrium toward a high-soul state experiences GDP gains on the order of 60 to 75 percent, corruption falling by two-thirds, and a shift from revenge to forgiveness.

5. We argue on theological, philosophical, and empirical grounds that Christianity is uniquely suited among the world's religions and ethical systems to supply the exogenous grace that the model requires. This is not a generic "religion is good for the economy" claim. It is a specific claim about the mechanism.

1.2 What this paper does not do

We do not claim that every Christian society has flourished or that every non-Christian society has failed. History is more complicated than any model. We do not claim that grace is the *only* determinant of economic outcomes. Physical capital, technology, geography, and institutions all matter. We do not claim that the soul state is directly measurable with current data, though we discuss proxies. And we do not claim that this paper will convince the skeptic. We claim only that the framework is internally consistent, that it nests the standard model, that it generates testable predictions, and that it explains patterns that the standard model cannot such as persistent corruption, the failure of institutional reform, and the power of forgiveness.

1.3 Plan of the paper

Section 2 reviews the related literature and identifies the gaps our model fills. Section 3 argues for the distinctiveness of Christianity as the source of the model's exogenous grace. Section 4 presents the formal model. Section 5 develops the four channels and seven propositions. Section 6 connects the model's two equilibria to Augustine's two cities. Section 7 presents numerical

simulations. Section 8 proves the nesting result. Section 9 addresses the secular economist directly. Section 10 concludes. The full Lagrangian derivation and extended proofs are in the Appendix.

2 Related Literature

Our paper sits at the intersection of five literatures: the economics of religion, social capital, institutions and growth, culture and economic outcomes, and behavioral economics. In each case, we build on important findings while addressing a specific limitation.

2.1 Economics of religion

Iannaccone (1998) provides the foundational survey, establishing that religious participation is amenable to economic analysis: agents choose denominations, invest time in religious activities, and respond to incentives within religious organizations. The economics of religion has since grown into a substantial field, treating religion as a club good, a form of insurance, or a source of network externalities.

Barro and McCleary (2003) take the field to macroeconomics. Using cross-country panel data, they find that religious *beliefs* (particularly belief in hell) positively predict economic growth, while church *attendance* has a negative effect conditional on beliefs, suggesting that the productive input is conviction, not time spent in pews. McCleary and Barro (2006) extend this analysis and propose a "religion market" framework where state regulation of religion affects both the supply and demand for spiritual goods.

Guiso et al. (2003) provide micro-level evidence that Christian denominations (particularly Protestantism) are associated with attitudes favorable to economic growth: trust, thrift, support for market competition, and the belief that poverty results from laziness rather than injustice. The last finding we report as theirs, not as ours. Effort matters, and a culture that honors work will outgrow one that despises it. But poverty is often imposed rather than chosen: extractive institutions manufacture it deliberately (Acemoglu and Robinson, 2012), for example apartheid. Our own model makes the point formally. In the low-S equilibrium the honest poor are impoverished by the dishonesty of others, by the rents extracted around them, not by any deficiency of their own effort. Where this paper diagnoses poverty, the diagnosis runs through the soul of a society, and especially the souls of its powerful, not through the supposed laziness of its victims.

What is missing. This literature establishes that religion *correlates* with economic outcomes but does not provide a formal model of the *mechanism*. Barro and McCleary treat religious beliefs as explanatory variables in a growth regression, but they do not model how beliefs enter the agent's optimization problem. Guiso, Sapienza, and Zingales document attitudinal correlations but do not explain *why* Christian denominations produce different attitudes, or why those attitudes persist across generations. Iannaccone models the choice to participate in religion but not the effect of religious transformation on economic behavior.

Our contribution. The soul state S enters the utility function and transforms preferences through four channels (patience, satiation, generosity, honesty). Grace ($\mathcal{G}_t$) is modeled as an

exogenous stochastic process, or the formal analogue of what Barro and McCleary measure as "belief." The model generates predictions that can be tested with the same data these authors use, but with a structural interpretation.

2.2 Social capital

Putnam (1993) demonstrated that civic engagement (measured by voter turnout, newspaper readership, and associational membership) predicts institutional quality across Italian regions, even after controlling for economic development. Putnam's insight launched the social capital literature. Knack and Keefer (1997) extended it to the cross-country setting, showing that generalized trust (measured by the World Values Survey question "Can most people be trusted?") predicts GDP growth after controlling for initial income, education, and investment rates. Fukuyama (1995) made the popular case that trust is the foundation of prosperity, arguing that "low-trust" societies (his examples: southern Italy, China, France) suffer from small firm size, weak institutions, and the costs of formal contracting.

What is missing. The social capital literature takes trust as a primitive; a cultural endowment that varies across societies but whose *origin* is left unexplained. Putnam traces Italian civic capital to medieval communes, but this pushes the question back: why did some regions develop communes and others did not? Knack and Keefer show that trust predicts growth but cannot explain why trust erodes in some periods and revives in others. Most critically, social capital is modeled as a *stock* that enters the production function or the constraint set, but it does not *transform preferences*. An agent in a high-trust environment faces lower transaction costs (a constraint effect) but still maximizes the same utility function. If the environment changes, behavior reverts. This cannot explain the deep, persistent behavioral changes associated with genuine moral transformation.

Our contribution. We model the soul state as a *preference transformer*, in addition to a constraint shifter. Higher S does not merely reduce transaction costs. It changes what the agent wants, how patient the agent is, and whether the agent can cheat. This is a deeper form of social capital that persists even when the external environment deteriorates, because the change is internal. Furthermore, we provide an origin story for social capital: it derives from the soul state, which itself depends on grace and spiritual investment. Social capital, in our framework, is *endogenous to the soul*. It is the observable manifestation of an unobservable spiritual transformation.

2.3 Institutions and growth

North (1990) established the modern framework: institutions are the "rules of the game" (formal constraints like laws and constitutions, and informal constraints like norms and conventions) that structure economic exchange. Good institutions reduce transaction costs, protect property rights, and enable impersonal trade. Acemoglu and Robinson (2012) argue that the fundamental cause of differences in prosperity is *inclusive* versus *extractive* institutions, tracing institutional origins to historical contingencies (colonialism, the Black Death, the Glorious Revolution).

The institutions literature has been enormously productive. But it faces an increasingly uncomfortable empirical puzzle: *institutional reform does not reliably produce institutional quality.*

Countries adopt constitutions, create anti-corruption agencies, establish independent judiciaries and central banks and yet corruption persists, rule of law decays, and growth stagnates. South Africa is the canonical example: one of the most admired constitutions in the world, robust Chapter 9 institutions, an independent Reserve Bank, and yet have governance indicators that have deteriorated relentlessly since 2008. The institutions-as-rules framework cannot explain why the same rules produce different outcomes in different societies.

What is missing. North himself acknowledged the gap. His informal constraints (norms, conventions, codes of conduct) do much of the explanatory work, but he provides no theory of where they come from or how they change. Acemoglu and Robinson's historical contingency argument is powerful for explaining long-run divergence but offers no prescription for a society trapped in extractive equilibrium today. "Change your institutions" is advice that presupposes the existence of agents willing and able to design, implement, and sustain good institutions. Where do such agents come from?

Our contribution. We provide the missing micro-foundation for institutions. Good institutions require good agents; agents who are patient enough to invest in public goods, honest enough to resist corruption, generous enough to share power, and forgiving enough to sustain cooperation after conflict. These are the characteristics that rise with the soul state S. The model predicts that institutional reform without spiritual transformation will fail (the institution operates in the low-S equilibrium and is captured by low-S agents), while spiritual transformation without institutional reform can still improve outcomes (high-S agents generate trust, honesty, and cooperation even in weak institutional environments).

2.4 Culture and economic outcomes

Tabellini (2010) shows that cultural traits such as trust and respect for others, measured at the regional level in Europe predict economic development even after controlling for country fixed effects and historical institutions. He argues that culture is transmitted intergenerationally and co-evolves with institutions. Alesina and Giuliano (2015) survey the broader literature on culture and economics, documenting correlations between cultural variables (individualism, family ties, generalized morality) and outcomes (growth, innovation, redistribution). They emphasize that culture is slow-moving, persisting over centuries, and that it interacts with institutions in complex ways. The most ambitious recent contribution is Henrich (2020), who traces the distinctive psychology of the West, its individualism, its analytic cast of mind, and above all its trust of strangers, to the medieval Western Church's dissolution of intensive kinship through prohibitions on cousin marriage and polygamy; Schulz et al. (2019) provide quantitative evidence that longer historical exposure to the Church predicts less cousin marriage and more impersonal prosociality today.

What is missing. The culture literature faces the same foundational problem as the social capital literature: culture is taken as given. Tabellini's intergenerational transmission model explains persistence but not change. If culture is transmitted from parents to children and shaped by historical institutions, then a society born into a "bad" culture is trapped with no mechanism for cultural renewal. The literature documents correlations between cultural traits and economic outcomes but provides no formal model of how cultural traits enter the agent's

decision problem, no theory of cultural transformation, and no explanation for episodes of rapid cultural change (post-war Germany, post-genocide Rwanda, the East Asian tigers).

Our contribution. Our model provides a formal mechanism for cultural change. The soul state S is the individual-level analogue of what the culture literature measures as "trust," "respect," or "generalized morality." It is transmitted intergenerationally (through family devotional investment and the social multiplier ξ), it co-evolves with institutions (through the feedback between S and governance quality), and it is slow-moving in normal times (the low and high equilibria are locally stable). But importantly it can change rapidly when grace intervenes. The exogenous grace shock ε_t^G is the model's explanation for episodes of rapid cultural transformation that the intergenerational-transmission framework cannot account for. The Wesleyan revival, the post-war Rwandan reconciliation, and the rapid Christianization of South Korea that preceded its economic take-off are examples of these grace shocks in our framework.

2.5 Behavioral and preference foundations

A growing body of work in behavioral economics has, independently and without invoking the soul, established micro-foundations for each of our four channels. These papers are important because they demonstrate that our model's mechanisms are not ad hoc theological impositions but formalize behavioral regularities that mainstream economists have already documented.

Honesty. Mazar et al. (2008) develop a theory of "self-concept maintenance" in which individuals balance the financial benefits of cheating against the psychological cost of revising their moral self-image. Agents cheat only up to the point where dishonesty would force a negative update of their self-concept. This maps directly onto our honesty constraint $\theta_t \leq \bar{\theta}(1 - S_t)$: the soul state S determines the agent's moral self-concept, which in turn sets the internal ceiling on dishonesty. Mazar, Amir, and Ariely provide the behavioral evidence; we provide the state variable that moves the ceiling.

Identity and moral assets. Bénabou and Tirole (2011) formalize how agents invest in their identity because they care about "who they are." In their model, past moral actions serve as signals of the agent's type, and agents invest in moral behavior partly to maintain a favorable self-image that helps resist future temptation. This is the closest secular analogue to our soul as a state variable: identity is an accumulated asset that reshapes future behavior. A reader may reasonably ask whether the soul is simply identity capital under a different name. It is not - the distinction is substantive on two counts. *First, the locus of the change.* Bénabou and Tirole keep the utility function fixed: identity is a belief about one's own type, managed through self-signaling, and it operates by adding a self-image term to an otherwise standard objective. The soul, by contrast, *transforms the utility function itself*: it moves the discount factor β, the curvature σ, the generosity weight ψ, and the honesty ceiling $\bar{\theta}(1 - S)$. Identity capital changes what an agent will pay to protect a self-image; the soul changes what the agent wants. *Second, and economically more consequential, the source.* In their framework identity investment is endogenous and self-referential: the agent builds moral capital, for the self, and can in principle bootstrap. In ours, the deepest increments to the soul come from outside the agent entirely, through grace, which is exogenous and cannot be self-generated. If moral capital is self-produced, a low-morality agent can invest his way out; if the decisive input is exogenous

grace, the low equilibrium is *not* escapable by individual effort alone. The exogeneity is precisely what creates the poverty trap of the soul that the rest of this paper turns on (Section 3.3). We should be candid that this is an empirical wager, not a definition, and it is a point of contrast between the two models. If moral capital were self-produced, a society with the institutional scaffolding to reward it, such as courts, audit bodies, and an independent central bank, should be able to bootstrap its way toward the high equilibrium. The model predicts it cannot. The persistent failure of institutional reform in well-designed environments (Section 9) is the evidence on which we stake the claim. Were self-improvement to lift such societies reliably out of the trap, the exogeneity assumption would be wrong.

Endogenous time preferences. Becker and Mulligan (1997) developed the foundational model showing how agents expend resources to "produce" patience, thereby raising their future discount factor β. Their framework demonstrates that differences in time preference can be endogenous to investment decisions, leading to divergent steady states in wealth and savings; a result that parallels our Proposition 2. Our patience channel $\beta(S)$ extends their insight by identifying a specific source of patience investment (devotional practice) and a specific mechanism for its transformation (grace), rather than leaving the "imagination technology" unspecified.

Warm-glow giving. Andreoni (1990) established that individuals derive internal, psychological utility from the act of giving itself, a "warm glow," separate from the public good created. This is what our generosity weight $\psi(S)$ captures: the soul state determines how much intrinsic satisfaction the agent derives from giving. Andreoni's contribution is to show that warm-glow preferences rationalize observed charitable behavior that pure altruism cannot explain. Our contribution is to endogenize the warm glow: it is not a fixed taste parameter but a function of the soul state, which itself depends on grace.

Ethical agents and the economics of obligation. Closest in spirit to our project is the formal economics of ethical motivation. Feddersen and Sandroni (2006) model agents who are moved not only by self-interest but by what an agent "of their type" is *obligated* to do, and show that this resolves puzzles (such as costly voter turnout) that the purely instrumental model cannot. Their work demonstrates that ethical motivation can be formalized cleanly and taken to data. The instructive contrast is the *source* of the ethical rule: in Feddersen and Sandroni the operative obligation is *endogenous*, determined within the equilibrium by how others behave. In our model the deepest input, grace, is exogenous to the agent and to the equilibrium. The two approaches are complementary (theirs shows ethics can be modeled; ours asks where the binding ethical disposition ultimately comes from), but the difference in the locus of causation is the issue Section 3.5 takes up.

What is missing. Each of these papers establishes one channel in isolation. We try to connect these concepts and ask why some agents have stronger self-concept maintenance, higher identity investment, greater patience, or warmer glow than others. We provide a mechanism for *simultaneous* transformation across all four margins. The soul state S is the key variable that moves honesty constraints, discount factors, generosity weights, and satiation parameters together. The behavioral literature has, unwittingly, been documenting the economic shadow of the soul, one channel at a time.

2.6 Summary: Positioning this paper

Table 1 summarizes how our model relates to each literature.

Table 1: Positioning the trichotomous agent model in the literature

Literature	Key insight we build on	What our model adds
Economics of religion	Religious beliefs predict growth (Barro & McCleary)	Formal mechanism: S transforms β, σ, ψ, θ
Social capital	Trust predicts institutional quality (Putnam, Knack & Keefer)	Soul as *preference transformer*, not constraint shifter; origin of trust in grace
Institutions	Rules of the game matter (North, Acemoglu & Robinson)	Agents who sustain good institutions require high S; institutions without soul are hollow
Culture & growth	Cultural traits persist and predict outcomes (Tabellini, Alesina & Giuliano)	Mechanism for rapid cultural change: exogenous grace shocks produce S-curve transitions
Behavioral & preferences	Each channel has independent micro-evidence (Mazar et al., Bénabou & Tirole, Becker & Mulligan, Andreoni)	Soul as the *single* state variable that moves all four channels simultaneously; grace as the mechanism

3 Why Christianity Is Distinct

Having situated our contribution in the economics literature, we now turn to the theological claim that distinguishes this paper from generic "religion and economics" work. Christianity is *qualitatively different* from other religious and ethical systems in ways that matter for the model's mechanism. The distinction lies in claims that are unique to Christianity, or at least uniquely central to it.

Before the argument begins, the order of its elements must be set right, because the exposition that follows runs backwards. The true order starts with Christ. It is from Christ that grace is given: "grace and truth came through Jesus Christ" (John 1:17). Grace is not a freestanding force that the world's religions parameterize differently; it is a gift that flows from a person and from what that person did. The chain behind the gift runs further back. We need Christ because our nature is broken; the nature is broken because all have sinned (Romans 3:23); and sin is the will set against God's will (1 John 3:4). Christ is the ground, grace the derivative. The exposition nonetheless begins with grace, for a reason internal to the discipline: grace is the first element of the chain that economics can see. $\mathcal{G}_t$ is a term in a law of motion; Christ stands upstream of the model's boundary. So we enter at the boundary and work toward the source: Section 3.1 establishes the structure of grace and locates the world's traditions against it, and Section 3.2 reaches the ground and sets the order right.

3.1 Grace is a gift, not an achievement

The central mechanism in our model is grace ($\mathcal{G}_t$): an exogenous shock to the soul state that the agent cannot manufacture, earn, or control. It is a formalization of the Christian doctrine of *sola gratia*, or by grace alone.

In Christianity, salvation and moral transformation are gifts from God, received through faith, not achieved through human effort. "For it is by grace you have been saved, through faith—and this is not from yourselves, it is the gift of God—not by works, so that no one can boast" (Ephesians 2:8–9). The agent does not climb to God; God descends to the agent. This is the theological content of the exogeneity assumption.

Let the transformation of the soul S be driven either by the agent's own devotional and moral investment i_t^S (an *endogenous*, "self-power" channel) or by an unmerited exogenous term $\mathcal{G}_t$ (an "other-power" channel). Three cases are important. In the *pure-exogeneity* case, $\mathcal{G}_t$ arrives independently of i_t^S: receptivity itself is given rather than produced (strict monergism). In the *cooperative* case, $\mathcal{G}_t$ remains unearned and disproportionate to effort, yet i_t^S raises the rate at which it arrives (prevenient grace one cooperates with). In the *fully endogenous* case, $\mathcal{G}_t = 0$ and transformation is produced entirely by i_t^S (works). Our claim is not that the pure-exogeneity structure is unique to Christianity—it recurs, as a minority strand, elsewhere—but that it is *central and originating* in the Augustinian–Reformed tradition, where even the agent's receptivity is treated as gift, and that its distinctive theological *ground* (atonement and imputed righteousness) has no parallel, an argument Section 3.2 completes. Against this benchmark we locate the other traditions.

- **Judaism** is not one more entry in this list, and we do not treat it as one. It is the tradition from which Christianity received the judge, the law, and the very grammar of grace, and Section 3.2 gives it the separate treatment that closeness demands. Here we place only its rabbinic form in the taxonomy. Covenant grace is real and unmerited: God chose Israel "not because you were more in number than any other people," but because He loved them (Deuteronomy 7:7–8). Yet the moral transformation of the agent runs through repentance (*teshuvah*), which the tradition holds to be within the agent's power, divinely assisted rather than divinely given: "Return to me, and I will return to you" (Malachi 3:7); "one who comes to purify himself is helped" (b. Shabbat 104a). That is the *cooperative* case: $\mathcal{G}_t$ unearned and real, with its arrival rate rising in i_t^S.

- **Islam** emphasizes submission (*islām*) to God's law and the Five Pillars, with deeds weighed on the scales (*mīzān*) at judgment. It is important not to overstate this as salvation "earned": Islam holds a strong doctrine of divine mercy (*raḥma*) and bounty (*faḍl*)—the tradition insists none enters Paradise by deeds alone but only by God's mercy—and Ash'arī theology is strongly predestinarian. What is effort-driven is the moral *transformation of the agent* in this life, even if acquittal at judgment ultimately rests on mercy. On that variable the mechanism is endogenous: in our model, $\mathcal{G}_t = 0$ with a high return on i_t^S (devotional investment) and no exogenous breakthrough, with divine mercy entering at judgment rather than as an infused transformation of S.

- **Buddhism** (in its Theravāda and Zen forms) teaches that liberation is realized through

the Eightfold Path—right view, right intention, and so on—with the emphasis on self-effort (*jiriki*, "self-power"), mindfulness, and the reduction of craving (*taṇhā*) rather than desire as such. Here $\mathcal{G}_t = 0$, with an objective that minimizes craving rather than transforming it. The important exception is **Pure Land** Buddhism (Jōdo Shū, Jōdo Shinshū), whose *tariki* ("other-power") soteriology rests entirely on Amida's vow; in Shinran even entrusting faith (*shinjin*) is Amida's gift, not the practitioner's work. Pure Land thus instantiates the pure-exogeneity case within Buddhism, and we treat it accordingly rather than as a counterexample to the taxonomy.

- **Hinduism** offers multiple paths (*yogas*)—devotion, knowledge, action, meditation—each directed at the individual's effort toward *mokṣa*. In the effort-based strands (*jñāna* and *karma* yoga, and the "monkey-hold" [*markaṭa-nyāya*] reading of devotion) the devotee's love is the precondition of divine favor, and the mechanism is endogenous. But this must be qualified: the Teṅkalai school of Śrī Vaiṣṇavism develops a doctrine of "grace without a cause" (*nirhetuka-kṛpā*)—the "cat-hold" (*mārjāra-nyāya*), in which the kitten is carried passively—in which grace is entirely unearned and prior to the devotee's love, anchored in Bhagavad-Gītā 18.66. This southern strand is close to the pure-exogeneity case, and we note it as the bhakti tradition's nearest analogue to unmerited grace rather than claiming its absence.

- **Secular humanism** rejects the supernatural and locates moral progress in reason, education, and institutional design. In our model this corresponds to $S = 0$: the soul is assumed away and the entire burden of improvement falls on the mind M and external institutions. A central question of this paper is whether that framework can account for persistent dysfunction in institutional environments that are, by the standard design criteria, well-constructed—a question we take up in Section 9 rather than settle by assumption here.

The economic consequence of exogenous grace is profound. If moral transformation is endogenous, or if it depends entirely on the agent's investment, then a society trapped in the low equilibrium *cannot escape by its own effort*. The shadow value of spiritual investment (μ^S) is low when S is low, the social multiplier works against the individual, and the rational response is to not invest. This is the poverty trap of the soul. Only an exogenous shock can break the cycle. Christianity claims that such a shock is available, always ($\bar{G} > 0$), and sometimes in extraordinary measure ($\varepsilon_t^G > 0$ which may include revival, awakening, and conversion).

3.2 The ground of grace: corrupted nature, satisfied justice, imputed righteousness

The taxonomy above concedes something. Pure Land Buddhism and the Teṅkalai school both instantiate the pure-exogeneity case, so exogeneity alone cannot carry the distinctiveness claim. Nor can the observation that other religions lack Christ: every tradition trivially requires its own machinery, and a claim of the form "Christianity is unique because it requires Christ" is circular until we say *why* Christ is required. The defensible claim relocates the uniqueness from the *structure* of grace, which is shared, to its *ground*: the problem grace must solve, and the

mechanism by which it solves it. Christianity diagnoses two problems where the other traditions see at most one.

Where do the two problems come from? The corruption we are about to assert is not an axiom; it is the conclusion of a chain that starts where Genesis starts, with a command and a will. Sin is the will set against God's will: "sin is lawlessness" (1 John 3:4). The command in Eden tested the will, and the test was pure because the prohibition was arbitrary in content: nothing about the fruit was harmful in itself, and Augustine's point is that the evil was never in the fruit but in the disobedience (Augustine, 426). A command with no independent rationale isolates the only thing at stake, whether the creature's will submits to the Creator's (Genesis 2:17). The will failed by inverting, preferring self to God: Augustine's *amor sui*, the seed of the *libido dominandi*. The inversion is inherited: "sin entered the world through one man" (Romans 5:12), and "all have sinned" (Romans 3:23). What a broken will cannot do is re-will itself. The two problems that follow are this chain restated as requirements: the inverted will needs power from outside, and the guilt of the inversion needs satisfaction.

The first is a problem of **power**. In the Christian diagnosis human nature is itself corrupted: "the mind governed by the flesh is hostile to God" (Romans 8:7); we are "by nature children of wrath" (Ephesians 2:3). The enmity runs from the agent toward God, and the faculty that would undertake devotional investment i_t^S is the very faculty that is broken. This corruption is what forces the exogeneity. If the operator that converts devotional effort into transformation, $\phi^S(\cdot)$ in the soul equation (equation (7), specified in Section 4), is applied by a corrupted will, then no sequence of investments i_t^S can be trusted to reach the transformed state: the agent is repairing the instrument with the instrument that needs repair, and the fixed point cannot be reached from inside the system's own dynamics. This is the rigorous content of "grace cannot be manufactured," and it is what permits the pure-exogeneity case over the cooperative one. Notice, too, that the taxonomy of Section 3.1 is not arbitrary. Where a tradition affirms this diagnosis, other-power soteriology follows: Shinran's *bombu*, the foolish being of blind passions, is nearly Augustinian, and other-power follows from it in Japan as it followed in Hippo. Where a tradition denies the diagnosis, self-effort follows: in the Islamic doctrine of *fiṭra* every person is born rightly oriented toward God, forgetful rather than fallen, and self-powered moral effort is therefore coherent. Diagnosis determines mechanism.

The second is a problem of **guilt**. Beyond the corrupted faculty stands an objective liability: the agent is not merely unable but condemned, before a judge whose justice is not negotiable. Corruption calls for power from outside; guilt calls for satisfaction. The two problems demand different operators, where no amount of the first substitutes for the second. Regeneration repairs the faculty; it does not discharge the debt. A doctor may heal a thief's hands without repaying what the hands stole.

Once the two problems are separated, the comparative landscape becomes precise. Each exogenous-grace tradition fills the second point differently. Islam holds the justice problem in full force: the scales (*mīzān*) are real and the judge is holy. But it forecloses the Christian resolution explicitly: no bearer of burdens may bear the burden of another (Qur'an 6:164; 17:15), and the atoning crucifixion itself is denied (4:157). Justice is resolved by sovereign mercy joined to the agent's repentance: pardon, not payment. Pure Land Buddhism has transfer without

a judge. Amida's merit, accumulated over cosmic ages, is turned over (*ekō*) to the entrusting believer, with the believer's karmic burden carried. But karma is an impersonal debt; carrying it propitiates no one, because there is no holy person whose justice is at stake. Teṅkalai grace is spontaneous compassion overriding karma: neither satisfaction nor transfer. Christianity alone holds both. There is a justice that must be satisfied rather than waived, so that God may be "just and the justifier of the one who has faith in Jesus" (Romans 3:26), and there is a transfer that satisfies it, the imputation of an alien righteousness: "God made him who had no sin to be sin for us, so that in him we might become the righteousness of God" (2 Corinthians 5:21). The transfer completes the chain this section opened with. In one garden the first Adam's will broke; in another garden, Gethsemane, the second Adam's held: "not my will, but yours be done" (Luke 22:42). Paul draws the parallel: "through the disobedience of the one man the many were made sinners, so also through the obedience of the one man the many will be made righteous" (Romans 5:19). What is imputed is the obedience the test required and the first will failed to give. The transaction is anchored in a datable public event, a historical atonement.

One tradition is deliberately absent from that landscape, because it does not belong among the rivals. Judaism is the root, not a branch: "theirs is the adoption to sonship; theirs the divine glory, the covenants, the receiving of the law" (Romans 9:4), and Paul warns the grafted branch against boasting over it, for "you do not support the root, but the root supports you" (Romans 11:18). Every element of the Christian ground is Hebrew before it is Christian. The holy judge is Israel's God. The election is unmerited (Deuteronomy 7:7–8). Abraham "believed the LORD, and he credited it to him as righteousness" (Genesis 15:6), the verse on which Paul builds the doctrine of justification (Romans 4:3). Substitution is already the logic of the sacrificial system, blood given on the altar "to make atonement for your souls" (Leviticus 17:11), and of the servant on whom "the iniquity of us all" is laid (Isaiah 53:6). On our reading, biblical Israel holds the entire structure of Table 2 in promissory form, and the disagreement between the two faiths sits exactly where they have always placed it: whether the Messiah has come and the atonement been made.

What belongs in the comparative table is therefore not Israel but the tradition that rebuilt Judaism after the Temple fell in 70 AD and the sacrificial system ceased. Rabbinic Judaism is a different configuration, and it answers the two problems differently. The power problem it declines: in place of a corrupted nature stand two inclinations, the *yetzer ha-ra* and the *yetzer ha-tov*, and the agent is charged to master the first ("sin is crouching at your door; it desires to have you, but you must rule over it," Genesis 4:7, with Deuteronomy 30:14 read the same way: "the word is very near you"). Diagnosis determines mechanism here as everywhere: with no broken faculty, repentance within the agent's power is coherent, and *teshuvah*, joined to prayer and charity, replaces the altar as the means of atonement. The guilt problem it holds in full force, before the same holy judge, but the verdict is annual: written on Rosh Hashanah, sealed on Yom Kippur, and reopened the next year. There is transfer language at the margin, the merit of the fathers (*zekhut avot*) invoked in the liturgy, but no imputation that settles the agent's standing once for all. Table 2 summarizes.

Why should an economist care which tradition fills the second slot? Because the second slot determines whether the agent's standing is settled during mortal life, and the timing of the

Table 2: The ground of grace: how traditions fill the two slots

	Nature corrupted (power problem)	**Justice to be satisfied (guilt problem)**	**Alien righteousness transferred**	**Standing settled in this life**
Christianity (Augustinian–Reformed)	Yes (original sin)	Yes (Rom. 3:26)	Yes (imputation)	Yes (justification)
Rabbinic Judaism (post-Temple)	No (two inclinations)	Yes (holy judge)	No (*zekhut avot* an echo)	No (sealed annually)
Pure Land Buddhism (Shinshū)	Yes (*bombu*)	No (impersonal karma)	Yes (*ekō*, merit)	Yes (*shinjin*)
Teṅkalai Śrī Vaiṣṇavism	Qualified	No (karma overridden)	No	Yes (the surrendered)
Islam	No (*fiṭra*)	Yes (*mīzān*)	No (Qur'an 6:164)	No (awaits judgment)
Works-piety (any tradition)	Denied in practice	Yes	No (self-payment)	No (earned, hence uncertain)

verdict changes the agent's problem. Consider three commitments. *Justice*: sin is an objective liability before a holy judge, to be satisfied rather than waived. *Inability*: the corrupted agent can neither repair the faculty nor discharge the liability; this is the first problem extended to payment. *Assurance*: the agent's standing can be settled now, once for all, within the mortal horizon. Hold all three and the conclusion is forced: someone else must pay, and the payment must be counted as the agent's own. Substitution and imputation are not one doctrinal option among several; they are the unique solution to the system. Every rival tradition relaxes at least one commitment instead. Works-piety, inside and outside Christianity, relaxes inability: the agent pays down the debt himself. Islam relaxes assurance: the verdict waits for the scales. Rabbinic Judaism relaxes both of those at once: *teshuvah* lies within the agent's power, and the verdict is sealed annually rather than once for all. Pure Land and Teṅkalai relax justice: the debt is carried or overridden, not paid.

Each relaxation has an economic footprint, with the model pricing all three. Relaxing *inability* produces works-based religion, whose dynamics Section 3.8 dissects: giving and devotion become investments in spiritual standing, the generosity prediction is contaminated, pride replaces gratitude, and the bimodal equilibria dissolve. Relaxing *assurance* produces verdict-pending devotion. Even with genuine grace in the first slot, the instrumental logic returns through the back door: devotional investment becomes a premium paid against an uncertain acquittal, and the works dynamics reappear in expectation even where grace is confessed. History shows the same margin operating inside Christendom. Where pardon was priced and repeatable, in the medieval confessional and the indulgence trade, the subjective cost of cheating fell; Blum and Dudley (2001) document that Protestants who could no longer wipe the slate clean faced a higher cost of cheating strangers, and honest dealing spread through wider trading networks. Rabbinic Judaism states the timing of its verdict with liturgical exactness: on Rosh Hashanah it is written and on Yom Kippur it is sealed, and "repentance, prayer, and charity avert the severity of the decree" (*Unetaneh Tokef*). The verdict-pending margin is not a hypothetical

there; it is a calendar.

Relaxing *justice* is the subtlest case. Shinshū faith preserves assurance and gratitude, and it is therefore the structural analogue of the Protestant ethic. The model predicts a partial economic footprint wherever the structure partially holds, and the record obliges: Bellah (1957) documents that Jōdo Shinshū supplied Tokugawa Japan's merchant class with an inner-worldly discipline of gratitude, the nearest Japanese analogue of the Protestant ethic. We count this as out-of-sample confirmation, not as a counterexample. What the deflated second slot cannot supply, on our reading, is the anchor of the honesty channel. In the Christian structure the moral law was satisfied at infinite cost, not suspended, so it retains its full objective authority for the pardoned agent: sin no longer moves the verdict, but it still grieves a holy person, and the corrosion term $-\rho\theta_t$ is that grief written into the law of motion. Where the debt is carried or overridden rather than paid, the continuing seriousness of sin rests on psychology, on gratitude and shame, rather than on a standing forensic fact. We state this as the model's wager rather than as settled comparative religion: traditions that relax justice should show the gratitude economics but a weaker anchoring of the honesty margin.

Rabbinic Judaism, on this same pricing, should show a real economic footprint with a particular shape, and it does. The configuration is strong on law, study, and covenant community, and the model reads each strength onto its own channel. The economic history is emphatic on the first two: Botticini and Eckstein (2012) show that when the Temple fell and the religion re-centered on the reading of Torah, the requirement of universal male literacy built human capital centuries before commerce paid for it, and later selected and equipped Jewish communities for skilled urban occupations. In the model's language that is the mind channel: a tradition of law and study writes its signature in M, and the Jewish economic record is, on our reading, the clearest historical confirmation the mind channel has. Giving is anchored differently than in our generosity channel: *tzedakah* is cognate with justice, an obligation owed rather than a warm glow enjoyed, closer to a constraint on the feasible set than to $\psi(S)$ in the utility function. And honesty toward the wronged party is built into the atonement itself: Yom Kippur atones for sins against God, but "for transgressions between a person and his fellow, the Day of Atonement does not atone until he has appeased his fellow" (Mishnah Yoma 8:9), which writes restitution into the religious year. So the model does not need to explain away Jewish prosperity, any more than it needed to explain away Japanese honesty (Section 9.2); it predicts the composition: a large footprint through the mind and through law-anchored conduct, with this paper's specific claim, preferences transformed by grace rather than conduct shaped by law, as the margin on which the two faiths genuinely differ. That too is the model's wager, offered in the same breath as the Shinshū case, not settled comparative religion.

The objection to a settled verdict is as old as the doctrine: "Shall we go on sinning so that grace may increase?" (Romans 6:1). The economist will recognize the shape immediately: moral hazard. If standing no longer depends on works, why not cheat to the moral limit? The model's answer is the theology's answer. Imputation never travels alone. The same effectual grace that settles the standing regenerates the faculty; the two are distinct in kind but inseparable in the order of salvation (see the Remark on justification and sanctification in Section 4). The regenerated agent does not exploit the settled verdict because the utility function has changed:

she no longer wants what the license would buy. And sin remains costly even though the verdict is fixed, because corrosion operates on the soul, not on the standing: $-\rho\theta_t$ hardens the heart whether or not the heart is justified. Imputation without regeneration would indeed be moral hazard. Regeneration without imputation would be anxious striving. The conjunction is the design.

The conjunction also reframes Weber. Weber's burgher worked from anxiety, scanning his ledgers for signs of election (Weber, 1905). But anxiety is the phenomenology of imputation disbelieved: it is what assurance feels like when it fails. Taken at its word, the doctrine supplies what Weber's mechanism lacks. If righteousness is imputed, work can earn nothing before God, and is thereby set free to serve the neighbor; Luther's doctrine of vocation (*Beruf*) is that liberation given an address. The engine of the ethic is not fear seeking evidence but gratitude seeking expression, and it points the four channels into the world rather than out of it: stewardship rather than withdrawal, honest dealing in the market rather than ascetic retreat, the patient building of institutions rather than the anxious hoarding of signs. Section 3.6 develops the relation to Weber in full; here we note only that the mechanism we formalize is not a rival psychology bolted onto his sociology but the doctrine's own logic, restored.

We can now state the distinctiveness claim in the model's language. One could decompose grace into a regenerative and a forensic component, $\mathcal{G}_t = \mathcal{G}_t^R + \mathcal{G}_t^F$, but the notation would mislead: the forensic component is not a flow into S. Let $J \in \{0, 1\}$ denote the agent's standing before God. Imputation sets $J = 1$ once for all at the moment of effectual grace, and J then appears in no law of motion and no period utility. Its entire model content is a set of exclusion restrictions. Because the verdict is settled, the soul equation (7) contains no works term κg_t (Section 3.8); the horizon carries no terminal verdict payoff V_{T+1} (the Remark on the horizon in Section 4); and the honesty condition prices no purchasable pardon. In a verdict-pending system J_T is instead a random variable whose distribution the agent can influence, and every excluded term reappears: giving earns, the horizon pays, pardon has a price. The ground of grace is therefore, formally, an identifying restriction. Traditions that fill the second slot differently do not imply a different rhetoric around the same model. They imply a different agent's problem, with different first-order conditions and different observable behavior. That is what makes the theological claim testable, and it is why the paper's Christianity is the source of its functional forms.

3.3 Why not synergistic grace?

An alternative specification, natural to the Catholic and Arminian traditions, would model grace as *offered* universally but effective only when the agent cooperates through free will and devotional effort. Under such a formulation, $\mathcal{G}_t^{\text{eff}} = \mathcal{G}_t^{\text{offered}} \cdot h(d_t)$, where d_t is the agent's devotional cooperation and $h(0) = 0$: if the agent refuses to cooperate, offered grace yields no transformation.

We reject this specification on three grounds: theological, economic, and empirical.

Theologically, synergistic grace contradicts the scriptural witness on the priority and sufficiency of divine action in salvation. Christ teaches that no one can come to Him unless the Father draws them (John 6:44). Paul insists that salvation "does not depend on human desire

or effort, but on God's mercy" (Romans 9:16) and that God "has saved us and called us to a holy life; not because of anything we have done but because of his own purpose and grace" (2 Timothy 1:9). The work of transformation, once begun, is God's to complete: "He who began a good work in you will carry it on to completion until the day of Christ Jesus" (Philippians 1:6). These texts do not describe a cooperative project between God and the human will. They describe a sovereign rescue. The agent who has been genuinely transformed by effectual grace will indeed invest in devotion ($i^S > 0$), pray, worship, and serve. But these acts are *consequences* of regeneration, not conditions for it. To reverse the order, i.e., making human cooperation the trigger that activates divine grace, is to make salvation a joint venture and to relocate the decisive cause from God to the creature. This is the error that the Reformation's *sola gratia* was formulated to correct.

Economically, synergistic grace dissolves the paper's most distinctive contribution. If grace is activated by human devotional effort, then *any* religion with a devotional practice can produce spiritual transformation. Islam's Five Pillars, Buddhist meditation, Hindu *bhakti*, and Confucian ritual all involve substantial devotional investment ($d_t > 0$). The model would predict that all of these are equally capable of raising S and producing the high equilibrium. The paper's specific claim about Christianity (Section 3, Contribution 5) would collapse into a generic "religion is good for the economy" thesis: precisely the undifferentiated claim that we distinguish ourselves from in Section 2.

Moreover, synergistic grace eliminates the poverty trap of the soul. If agents can incrementally cooperate their way to higher S, the low equilibrium becomes escapable through collective moral effort. This sounds appealing but contradicts the central empirical puzzle the paper addresses: why do societies with excellent institutions (South Africa's Constitution, Chapter 9 bodies, independent Reserve Bank) nevertheless stagnate in corruption and dysfunction? If incremental cooperation could bootstrap a society out of the trap, institutional reform should have worked. It did not.

We should be precise about where the trap comes from in the model. The two stable equilibria are a property of the social complementarity $\xi\bar{S}$, not of grace as such: mathematically, a coordinated rise in devotional investment could shift the replenishment curve enough to move a society across the threshold. But in the low basin no individual has the incentive to begin, because the shadow value μ^S is low precisely when everyone else's $\bar{S}$ is low, so the coordinated bootstrap never starts. The exogeneity of grace is what supplies the push from outside this no-investment trap. The complementarity is what makes the two basins exist; grace is what moves a society between them.

Empirically, some examples often cited in favor of synergistic grace such as the secularization of Quebec in the Quiet Revolution, Ireland's rapid de-Christianization, are already explained by our model without modifying the grace specification. Spiritual depreciation δ^S operates continuously. When a prosperous society reduces devotional investment ($i^S \to 0$), the soul decays toward the grace floor $\bar{G}/\delta^S$. The mechanism is standard capital depreciation, not a withdrawal of synergistic cooperation. Rising wages increase the opportunity cost of devotional time (this is already captured in the equal-marginal-returns condition, equation 12), and the rational agent substitutes toward mind and body investment. Secularization is endogenous to

the *investment decision*, not to the *nature of grace*. The original model explains these patterns; the synergistic revision adds complexity without explanatory gain.

3.4 Transformation, not compliance: the model's mechanism

The model's central mechanical departure from the standard framework is that the soul transforms *preferences*, not just *constraints*. We state this carefully, because it is the part of the paper most easily misunderstood as a distinctiveness claim when it is in fact a statement about the *form* of the model.

Most ethical and religious systems, in their public face, operate on the constraint set. "Do not steal" is a constraint. "Pay your taxes" is a constraint. "Give 2.5% of your wealth to charity" (the Islamic *zakat*) is a constraint. These rules can improve outcomes within a given utility function; they are enforcement mechanisms that change the feasible set without changing what the agent wants. A model of compliance leaves preferences fixed and tightens the constraints.

Our model instead lets the state variable S *change what the agent wants*: the discount factor $\beta(S)$ rises (patience), the curvature $\sigma(S)$ rises (satiation), the generosity weight $\psi(S)$ rises, and the honesty ceiling $\bar{\theta}(1-S)$ tightens. The agent does not merely face tighter rules; the agent becomes a *different kind of agent*.

We must be candid about what this does and does not establish. Preference transformation is *not*, by itself, a uniquely Christian idea. Any model with a non-standard preference element (altruism, warm glow, endogenous time preference, the extinction of desire that Buddhism seeks) also "changes what the agent wants" relative to bare homo economicus. So this subsection does not, on its own, single out Christianity; it specifies the model's machinery. The genuinely distinctive Christian claims lie elsewhere, in the *source* of the transformation: that the decisive input is grace, which is exogenous, unearned, and cannot be self-generated (Sections 3.1–3.3). Scripture frames the transformation vividly: "if anyone is in Christ, the new creation has come" (2 Corinthians 5:17), the promise that you will *want* to be generous, that dishonesty will become abhorrent, that forgiveness will flow from having been forgiven. That framing motivates our reading of the channels as *sanctification*. But the load-bearing distinctiveness argument is the exogeneity of grace, not the bare fact that preferences move.

3.5 Moral dispositions are derivative, not foundational

A reader trained in game theory will object to any claim that "norms are derivative," and rightly so, so we must be precise about which norms we mean. A great many norms are genuinely *endogenous*: they are equilibria of coordination games. Which side of the road to drive on, which system of measurement to use, how to greet a stranger: all of these are conventions, multiple Nash equilibria selected by history. Game theory explains them perfectly well without any appeal to the soul. We claim nothing about them. Even *cooperative* behavior can be supported endogenously: the folk theorem shows that cooperation can be an equilibrium of a repeated game, and formal models of ethical action, such as Feddersen and Sandroni (2006), in which agents act on what their "type" is obligated to do, derive the operative obligation endogenously from how others behave.

Our claim is narrower and survives this objection. It is not about norm-following *behavior*, which game theory handles, but about the *intrinsic valuation of moral ends*: the disposition to value honesty, trust, and forgiveness non-instrumentally, as goods in themselves rather than as moves that happen to be incentive-compatible. *That* disposition, we argue, is derivative: it is the visible expression of an underlying soul state, not a free-standing primitive. The distinction is important. An agent who cooperates only because the repeated-game structure makes cooperation an equilibrium will defect the moment that structure breaks: when the game becomes one-shot, when monitoring fails, when the rents on offer are large enough. A high-S agent will not, because for her the valuation is internal. This is why the social-capital literature, which treats trust as a primitive (Putnam, 1993; Knack and Keefer, 1997), cannot say where trust comes from or why it survives the collapse of the institutions that once supported it.

Consider the empirical puzzle: why do some cultures sustain trust and others corruption? If the moral disposition is foundational, this question has no answer; it is just *there*, transmitted by families and institutions, and if you are born into a low-trust culture, that is your fate. But if it is derivative, if it flows from the soul state, which itself depends on grace and spiritual investment, then there is a causal structure beneath the surface, and the possibility of change.

The historical record supports this. The abolition of slavery, the development of hospitals, the creation of universities, the codification of human rights, the emergence of modern science all have deep roots in Christian thought and Christian communities (Stark, 2005; Woodberry, 2012). These are not coincidences. They are the institutional expression of a soul state shaped by grace.

A secular economist may respond: "Fine, but I can explain all of this with social capital and cultural evolution without invoking God." We address this objection directly in Section 9. For now, we note only that the secular account cannot explain why social capital emerges in some places and not others, why it decays in some periods and revives in others, or why institutional reform so frequently fails to produce the behavioral change it was designed to elicit. Our model can.

3.6 Relationship to Weber

The most famous argument connecting Christianity to economic outcomes is Max Weber's *Protestant Ethic and the Spirit of Capitalism* (Weber, 1905). Weber argued that Calvinist anxiety about predestination (the agonizing uncertainty over whether one was among the elect) drove Protestants to seek worldly "signs of election" through disciplined work, thrift, and capital accumulation. The unintended consequence was capitalism.

Our model relates to Weber's in three respects. We offer these as points of *complementarity and extension*. We note at the outset, too, that the Weber thesis is itself empirically disputed, and has been almost from the beginning: Samuelsson (1993) went through the historical record and found the link between Protestant doctrine and capitalist practice wanting, and modern work attributes much of the Protestant economic advantage to literacy and human capital rather than to a distinctive ethic (Becker and Woessmann, 2009). Not all of the evidence cuts against Weber, but the part that survives points away from his mechanism and toward ours. Blum and Dudley (2001) document rising wages in Protestant cities and falling wages in Catholic ones

between 1500 and 1750, and attribute the divergence not to a work ethic but to cooperation: a Protestant who defected could no longer wipe the slate clean in the confessional, the subjective cost of cheating a stranger rose, and honest dealing spread through larger trading networks. That is the honesty channel, not the anxiety channel, doing the work (and the cheap-absolution logic they identify is the same one we formalize as the corrosion of works-based religion in Section 3.8). Our argument does not stand or fall with how the debate resolves; we offer a different mechanism in any case.

First, the *mechanism* we emphasize is different. Weber's agent works hard because of *anxiety* (the fear of damnation, the desperate search for assurance). Our agent is transformed by *grace*; the experience of unmerited love that produces gratitude, not fear. Weber's agent accumulates capital to reassure himself. Our agent invests because a transformed heart values stewardship, sees the needs of others, and thinks long-term. The psychological and theological foundations differ: anxiety versus peace, earning versus receiving, performance versus gratitude. Whether Weber's Calvinist burgher is better described by anxiety or by grace is a question for the Weber specialists; we claim only that the grace mechanism is the one our model formalizes.

Second, the *scope* is broader. Weber focused on one channel: work ethic $\rightarrow$ capital accumulation. Our model identifies four simultaneous channels (patience, satiation, generosity, honesty) and derives seven propositions spanning corruption, savings, redistribution, TFP, forgiveness, unemployment, and multiple equilibria. The Weberian thesis is, in effect, a special case of our model restricted to the patience and satiation channels.

Third, the *applicability* is wider. Weber's thesis was historically specific. It explained the rise of capitalism in 16th–17th century Northern Europe. The comparative literature that followed already strained against that boundary: Eisenstadt (1968) asked whether the Protestant ethic, or some functional equivalent, could explain modernization beyond Europe, and argued that what mattered was the transformative capacity of a religious tradition rather than any single doctrine. Our model gives that conclusion formal content: the soul state S and its transformation by grace are not confined to a particular denomination or era. The model can speak to 21st-century South Africa, post-genocide Rwanda, or post-war South Korea, contexts far removed from Weber's Calvinist burghers.

We regard Weber as an important precursor, not a rival, and we do not claim to have produced a "better model of the Christian individual." That is a judgment for scholars of Weber and of the sociology of religion to make. His core insight, that theological commitments have economic consequences, was correct and remains the foundation we build on. Our contribution is to formalize a different mechanism (grace rather than anxiety) and a wider transformation (preferences across four channels, not effort alone), and to render both in a model that can be taken to data.

3.7 The dark side: When Christianity fails

We must address the obvious objection: Christianity has also produced Crusades, Inquisitions, slave-trading nations, and directly relevant to this paper, the apartheid theology of the Dutch Reformed Church in South Africa.

Our model explains this. Nominal Christianity such as church membership, doctrinal assent,

and cultural identification, is not the same as a high soul state. The model's predictions depend on S, not on religious affiliation. An agent who attends church but has $S \approx 0$ (no genuine transformation, no operative grace) behaves exactly like homo economicus: self-interested, impatient, dishonest up to the constraint. The apartheid theologians who blessed racial oppression were, in our model's terms, low-S agents who used the language of Christianity to sanctify the *libido dominandi*. Augustine would have recognized them immediately as citizens of the City of Man dressed in the garments of the City of God.

The prosperity gospel (the claim that God rewards faith with material wealth) is the contemporary version of this corruption. It inverts every channel in our model: it encourages *impatience* (demand your blessing now), *insatiability* (God wants you rich), *selfishness* (your faith, your reward), and *dishonesty* (fake miracles, exploitative tithing). In our framework, the prosperity gospel is not high-S Christianity. It is a $S \approx 0$ ideology that mimics the vocabulary of grace while denying its substance.

The testable prediction is clear: societies with high rates of *genuine* religious practice (measured by behavioral proxies such as volunteering, charitable giving, marital stability, honesty in experiments, rather than self-reported belief) should outperform societies with high rates of nominal affiliation. The emerging empirical literature on "believing versus belonging" supports this distinction (Barro and McCleary, 2003).

3.8 The corrosion of works-based religion

The prosperity gospel is the most visible corruption of Christianity, but there is a subtler and more widespread one: the belief that religious works *earn* spiritual standing. This is the works-based framework, and it does not merely fail to raise the soul state, but can actively corrode it.

Consider an alternative model in which charitable giving g_t enters the soul's law of motion. Take the soul equation we will specify formally in Section 4 (equation (7)) and add a works term: $S_{t+1} = (1-\delta^S)S_t + \phi^S(i_t^S) + \xi\bar{S}_t + \mathcal{G}_t - \rho\theta_t + \kappa g_t$, where $\kappa > 0$ rewards giving with spiritual capital. (The notation is defined in full in Section 4: the depreciation rate δ^S, the devotional production function ϕ^S, devotional investment i^S, the social multiplier $\xi\bar{S}$, grace $\mathcal{G}$, and the corrosion term $\rho\theta$; the reader unfamiliar with it may wish to glance ahead.) In such a framework, the first-order condition for giving changes. Our model yields $\psi(S_t)g_t^{-\sigma} = \tilde{\lambda}_t$: the agent gives because the soul makes giving intrinsically satisfying. The works-based model adds an investment return: the agent also gives because giving raises S, which raises β (more patience), raises σ (more satiation), and relaxes future constraints. Giving becomes an *investment with private returns*, not an expression of genuine other-regard.

This has three destructive consequences.

First, it contaminates the generosity prediction. In our model, $g = 0$ when $S = 0$ (Proposition 3): a soulless agent has no reason to give. In the works-based model, $g > 0$ even at $S \approx 0$, because the agent gives instrumentally to build spiritual capital and harvest the patience and satiation dividends. The model can no longer distinguish genuine generosity from strategic philanthropy. The Pharisee who gives ostentatiously to be seen (Matthew 6:2) and the widow who gives her last mite from a transformed heart (Mark 12:43–44) produce the same observable

g, but the Pharisee's soul is not growing. It is calcifying.

Second, works-based religion produces spiritual pride, or the conviction that "I have earned my standing before God through my moral effort." Pride is the quintessential low-S disposition. Augustine identified it as the root of the *libido dominandi*: the desire to dominate, to be above others, to be self-sufficient. In the language of our model, spiritual pride *lowers* ψ (the proud agent serves the self, not others), *lowers* σ (spiritual achievements become a form of insatiable consumption. The Pharisee always needs more visible righteousness), and *loosens* the honesty constraint (the self-righteous agent can justify dishonesty toward those deemed less worthy). The agent who believes she has earned her soul state is, paradoxically, further from genuine transformation than the agent who knows she has not. "God opposes the proud but shows favor to the humble" (James 4:6).

Third, works-based religion unravels the multiple equilibria result (Proposition 7). If g feeds back into S, the soul equation gains an additional convexity: agents at moderate S give, which raises S, which raises giving further. The unstable tipping point $\bar{S}^*$ softens into a smooth gradient, and the sharp "critical mass" threshold dissolves. The model loses its ability to explain why societies cluster at two distinct poles, high-trust flourishing or low-trust dysfunction, rather than distributing continuously along a spectrum. The empirical pattern is bimodal, not smooth (Knack and Keefer, 1997; Putnam, 1993). The works-based model cannot generate this.

The deeper point is theological. Grace transforms; works perform. The agent transformed by grace gives because giving has become natural to her; the utility function has changed. The agent performing works gives because giving is profitable to her; the utility function is unchanged, but the investment set has expanded. The first is sanctification. The second is sophisticated self-interest dressed in religious clothing. And the historical record, from the Pharisees to the medieval indulgence trade to the modern prosperity gospel, suggests that the second does not merely fail to produce the high equilibrium. It actively entrenches the low one, by giving agents the *appearance* of high S (visible piety, conspicuous charity) while preserving the *substance* of low S (pride, insatiability, instrumentalism). Christ's harshest words were reserved not for the openly sinful but for the religiously self-righteous (Matthew 23). The model should reflect this: works performed to earn standing do not build the soul. Grace does.

4 The Model

Consider an agent who lives for T periods and makes decisions about consumption, education, and devotion. The agent has three state variables:

- $B_t \in [0, 1]$: **Body**, physical health and vigor.
- $M_t \in [0, \infty)$: **Mind**, human capital (knowledge, skills, cognitive capacity).
- $S_t \in [0, 1]$: **Soul**, spiritual capital (moral orientation, character, relationship with God).

Remark (Labels and identification). A skeptic will object that S could be renamed X, Z, or "butterflies," and that calling it the soul settles nothing. True, and true of every theoretical variable: productivity is not made technological by the letter A. The model's content lies

not in the label but in the cross-equation restrictions imposed on S. It is a single persistent latent state that jointly shifts patience ($\beta'(S) > 0$), material contentment ($\sigma(S)$), generosity ($\psi'(S) > 0$), and the honesty ceiling ($\bar{\theta}(1-S)$), and whose law of motion responds to devotional investment, social reinforcement, grace, neglect, and moral corrosion (equation (7) below). A generic latent variable is observationally equivalent to the soul state only if it satisfies this entire set of restrictions at once; a variable that moved patience but not honesty, or honesty but not generosity, would be a different object. Empirically, then, the model tests for the existence of a common moral-preference state with this structure. Interpreting that state as the soul, renewed by grace, is the paper's substantive theological claim, with Section 3 arguing it on its own ground.

4.1 Preferences

The agent maximizes lifetime utility:

$$V_0 = \mathbb{E}_0 \sum_{t=0}^{T} D_t \, u(c_t, g_t;\, S_t), \qquad D_t \equiv \prod_{s=1}^{t} \beta(S_s), \quad D_0 \equiv 1 \tag{1}$$

where c_t is own consumption, g_t is giving (voluntary transfers to others), and S_t is the soul state. The cumulative discount factor D_t is the product of the one-period discount factors $\beta(S_s)$ realized along the soul path. We use the product form, rather than the more familiar β^t, precisely because β is not constant: the soul changes the agent's patience over the life cycle, and the product lets the effective discount rate move period by period with S. When S is constant the product collapses to the standard β^t. The soul enters the agent's problem in two ways:

1. It transforms the **discount factor**: $\beta(S_t) = \underline{\beta} + (\bar{\beta} - \underline{\beta})\, S_t$, where $0 < \underline{\beta} < \bar{\beta} < 1$.
2. It transforms the **shape** of the period utility function.

The period utility function is:

$$u(c_t, g_t;\, S_t) = \frac{c_t^{1-\sigma(S_t)}}{1-\sigma(S_t)} + \psi(S_t)\, \frac{g_t^{1-\sigma(S_t)}}{1-\sigma(S_t)} \tag{2}$$

where:

- $\sigma(S_t) = \underline{\sigma} + (\bar{\sigma} - \underline{\sigma})\, S_t$ is the curvature parameter (inverse elasticity of intertemporal substitution). Higher S means higher curvature: diminishing marginal utility sets in faster, and the agent is more easily satisfied with material goods (see Remark 4.1 for the precise content of this claim).
- $\psi(S_t) = \psi_0\, S_t^{\eta}$ with $\eta > 0$ is the weight on giving. When $S = 0$, giving yields no utility and the agent is purely self-interested.

The utility function as the will. One interpretive remark before the algebra proceeds. The utility function is the economist's formal representation of the will: it states what the agent wants and how badly. Section 3.2 defines sin as the will set against God's will, and S should be

read accordingly. It measures not the strength of the will but its *orientation*: the alignment of what the agent wants with what God wants. $S = 0$ is not an agent with unfortunate parameters; it is the economic portrait of the will that ate, fully inverted toward self: impatient, insatiable, ungenerous, and dishonest up to the constraint. That the soul transforms the utility function rather than the constraint set is therefore not a modeling convenience but the substance of the claim: salvation is a re-willing, not new information or new incentives.

On the functional forms. The soul state S is a latent index, not a directly measured quantity, which we have normalized to $[0, 1]$. Its scale is therefore free: any strictly monotone re-labelling of S leaves the model's content unchanged. We use this freedom to make $\beta(S)$ *linear* in S. This is a normalization, not an assumption, since we may always re-define the index so that patience rises linearly along it. Once that normalization is fixed, however, the linearity of $\sigma(S)$ and the power form $\psi(S) = \psi_0 S^\eta$ are genuine functional-form assumptions, because only one of the three channels can be made linear by choice of scale. We adopt them for parsimony: each is the simplest smooth, monotone form consistent with the channel's economic content (curvature rising with S, giving vanishing at $S = 0$), and each adds at most one free shape parameter (η, which governs how quickly generosity emerges). What the propositions actually require is *monotonicity* and the boundary behavior ($\psi(0) = 0$, σ and β bounded).

Remark (What the satiation channel does and does not deliver). σ is a loaded parameter: it governs relative risk aversion and the elasticity of intertemporal substitution, not only contentment. Two implications of a rising $\sigma(S)$ are clean and unit-free. First, the intertemporal elasticity $1/\sigma$ falls, so the high-S agent's consumption path is flatter: she has less appetite for shifting consumption toward the present. Second, the giving-to-consumption ratio $g/c = \psi(S)^{1/\sigma(S)}$ rises in σ whenever $\psi < 1$: satiation with own consumption tilts spending toward giving. The third, more intuitive reading, that higher σ lowers the marginal utility of consumption, is *not* unit-free: $\partial u_c/\partial\sigma = -c^{-\sigma}\ln c$ is negative only when $c > 1$, so it holds in units where consumption exceeds one (say, consumption measured in multiples of subsistence) and reverses below. Where a result needs the marginal-utility reading, we impose it directly as an assumption, $\partial^2 u/\partial c\,\partial S \le 0$ on the relevant domain, and flag it (Proposition 2). A unit-robust alternative is a contentment shifter, $u(c, S) = [(c + \chi S)^{1-\sigma} - 1]/(1 - \sigma)$ with $\chi > 0$, which delivers $u_{cS} < 0$ globally; we retain the $\sigma(S)$ form because it keeps the giving ratio in closed form and because none of the quantitative results turn on the difference.

4.2 Budget constraint

The agent earns legitimate income from human capital applied to production, and can borrow and lend at an exogenous interest rate. In addition, the agent has access to a pool of appropriable rents $\Pi_t > 0$ (public resources, positions of trust, contractual opportunities where cheating is feasible) and can divert a fraction θ_t of these rents to private use:

$$c_t + g_t + i_t^B + i_t^M + i_t^S + a_{t+1} = w_t\, M_t\, B_t + \theta_t\, \Pi_t + (1 + r_t)\, a_t \tag{3}$$

where w_t is the wage per effective unit of labor, $M_t B_t$ is effective labor (human capital weighted by health), i_t^B, i_t^M, i_t^S are investments in body, mind, and soul respectively, and a_{t+1} is a one-period financial asset carrying the gross return $1 + r_{t+1}$ into the next period. The asset is the ordinary vehicle for moving resources across time; without it the budget would bind period by period and saving would not even be defined, so the consumption Euler equation below could not be derived from the stated problem. We take r_t as given (a small-open-economy or partial-equilibrium reading), set a_0 given, and impose the terminal condition $a_{T+1} \geq 0$: the agent cannot die in debt.

The rent pool Π_t represents resources that are vulnerable to appropriation: government budgets, aid flows, tax revenues, contractual obligations, employer assets, communal property. In a developing economy with weak institutions, Π_t is large. The critical feature is that corruption *benefits the individual* ($\theta_t \Pi_t$ enters the agent's income) while *harming aggregate output* (resources diverted from productive use reduce effective labor; see Section 5). This wedge between private gain and social cost is the fundamental tension that the honesty constraint must resolve.

Remark (Where do rents come from?). We treat Π_t as exogenous to the individual agent, determined by the institutional environment. In a richer model, Π_t could be endogenized: weaker governance (itself a function of $\bar{S}$) creates larger rent pools, generating a feedback loop where low soul states produce weak institutions that produce more temptation. For the present analysis, taking Π_t as given is sufficient to establish the mechanism.

4.3 The honesty constraint

The agent faces an internal moral constraint on dishonesty:

$$0 \leq \theta_t \leq \bar{\theta}\,(1 - S_t) \tag{4}$$

where $\bar{\theta} \in (0, 1)$ is the maximum feasible appropriation rate. Both bounds are economically live: the upper bound is conscience, and the lower bound binds when the agent freely chooses full honesty, which the corrosion mechanism below makes possible; the Kuhn–Tucker conditions in Appendix A carry a multiplier for each. When $S_t = 1$ (fully Christ-centered), the constraint forces $\theta_t = 0$ which implies that the agent *cannot* cheat because internal moral conviction prevents it. When $S_t = 0$, the constraint is slack and the agent appropriates up to $\bar{\theta}$ of available rents.

This is not an external enforcement mechanism. There are no police, no courts, no regulators in this constraint. It is a formalization of conscience; an internal limit that tightens as the soul grows. The agent *wants* to cheat (it raises income by $\theta_t \Pi_t$) but the soul prevents it.

4.4 Laws of motion

Body. Physical health depreciates with age and can be maintained through health investment:

$$B_{t+1} = (1 - \delta_t^B)\, B_t + \phi^B(i_t^B) \tag{5}$$

where δ_t^B increases with t (aging) and $\phi^B(\cdot)$ is concave. The terminal condition is $B_T = 0$ (death).

Mind. Human capital accumulates following Ben-Porath (1967) with depreciation:

$$M_{t+1} = (1 - \delta^M)\, M_t + \phi^M(i_t^M, M_t) \tag{6}$$

where $\phi^M(i_t^M, M_t) = A^M\, (i_t^M)^\gamma\, M_t^{1-\gamma}$ is the Ben-Porath production function. Existing human capital aids the acquisition of new human capital.

Soul. The soul evolves according to:

$$S_{t+1} = \mathcal{P}_{[0,1]}\big[(1 - \delta^S)\, S_t + \phi^S(i_t^S) + \xi\, \bar{S}_t + \mathcal{G}_t - \rho\, \theta_t\big] \tag{7}$$

where $\mathcal{P}_{[0,1]}[x] \equiv \min\{1, \max\{0, x\}\}$ projects the transition onto the unit interval. The projection is required for the state space to be well defined: a large grace realization could otherwise carry S above one, and heavy corrosion at a low soul state could carry it below zero. It never binds on the interior paths we study (the simulations impose it explicitly), and every derivative-based statement below refers to interior S, where the bracketed law applies as written. Here δ^S is the spiritual depreciation rate (decay from neglect), $\phi^S(i_t^S)$ is a concave production function for spiritual growth through devotional investment (prayer, worship, fellowship, study of Scripture), $\xi\, \bar{S}_t$ is a social multiplier capturing peer effects from the average soul state $\bar{S}_t$ in the community (it is easier to grow spiritually when surrounded by others doing the same), and $\mathcal{G}_t$ is grace. The parameter $\xi > 0$ governs the strength of this complementarity and is the key source of multiple equilibria (Proposition 7).

The final term, $-\rho\, \theta_t$ with $\rho > 0$, is the *corrosion of the soul by dishonesty*: cheating today does not merely transfer rents, it hardens the heart and lowers tomorrow's soul state. This term creates a genuine intertemporal cost of corruption that the agent internalizes, and it restores an interior honesty choice: the agent may decline to cheat even when the moral constraint (4) would still permit it, because the future spiritual cost outweighs the immediate gain (see Section 5). The specification embeds a deliberate *asymmetry*, consistent with the argument of Section 3.8: grace ($\mathcal{G}_t$) and the believer's devotional investment ($\phi^S(i_t^S)$) build the soul, while neglect (δ^S) erodes it and sin ($-\rho\theta_t$) corrodes it. What pointedly does *not* build it is *charitable giving*: there is, by design, no term in g_t. Giving is the fruit of a transformed soul, never a lever for manufacturing one, the works-righteousness error dissected in Section 3.8.

Remark (Justification and sanctification). A theologically careful reader will want this stated precisely. The state variable S tracks *sanctification*, the progressive transformation of character and desire, not *justification*, the once-for-all forensic verdict by which the believer is declared righteous. The model does not attempt to represent justification, which in the Reformation traditions this paper draws on is received through faith and not forfeited.

One subtlety deserves its own word, since it is easy to miss. The effectual-grace shock ε_t^G (Section 4.5) might look like justification under another name: it too is sovereign, particular, and grounded in election. But it is not. The shock ε_t^G is *regenerative*: it changes what the agent *is*, the soul state itself, and with it the agent's desires and the very faith by which justification is received. Justification is *forensic*: a change not in the agent's character but in the agent's

standing before God. In the order of salvation the two are inseparable (the regenerated person is the justified person, and grace, not foreseen faith, is the root of both), yet they remain distinct in kind. The model carries the renovative thread, regeneration issuing in sanctification. The forensic verdict enters differently: not as a term in any equation but as a set of restrictions on the model's form (Section 3.2); because the verdict is settled, the soul equation carries no works term and the agent's problem no terminal payoff.

Sanctification is different in kind: it is commanded. "Work out your salvation with fear and trembling," Scripture says, at once adding, "for it is God who works in you" (Philippians 2:12–13). The model is built to honor both clauses: $\phi^S(i_t^S)$ is the believer's working-out through the ordinary means of grace (prayer, worship, study, fellowship), while $\mathcal{G}_t$ is God working within. And, consistent with Scripture's warnings against grieving the Spirit and its language of fatherly discipline and correction, the model lets sanctification *regress*: the soul decays through neglect (δ^S) and corrodes through sin ($-\rho\theta_t$). Butner (2024) develops the same Philippians text into a theology of markets: participation in economic life is itself morally formative, building or deforming the character of the participant. The model agrees on the deforming side, since the corrosion term $-\rho\theta_t$ is market conduct writing itself back onto the soul, and it is deliberately narrower on the building side: honest work and giving are fruits of the soul here, not levers for manufacturing it (Section 3.8). What the model denies is narrower than it may first appear: not that the believer labors, but that the soul can be *earned*, that a tally of works could oblige God to supply it. Grace remains the engine even as the believer works.

4.5 Grace

Grace follows:

$$\mathcal{G}_t = \bar{G} + \varepsilon_t^G, \quad \varepsilon_t^G \sim F(\cdot), \quad \varepsilon_t^G \geq 0 \tag{8}$$

The two components correspond to a theological distinction. $\bar{G} > 0$ is *common grace*: God's persistent kindness toward all creation such as sustaining life, restraining evil, granting the rain to fall on the just and the unjust (Matthew 5:45). Common grace is universal and never withdrawn. It is the reason that even thoroughly secular societies retain a residual moral sense, that cooperation does not collapse entirely, and that beauty persists in a fallen world. The stochastic term ε_t^G represents *effectual grace*: the specific, transformative work of the Holy Spirit that regenerates the heart and grants the gift of faith. Effectual grace produces conversion, revival, and awakening. It is particular, given to those whom God has chosen before the foundation of the world (Ephesians 1:4), and it is irresistible in the sense that when God purposes to transform a soul, the transformation occurs.

Grace is the model's most original element. Four features deserve emphasis:

1. **Exogeneity.** The agent cannot manufacture, earn, or control grace; neither common nor effectual. Even the faith to believe is itself a gift (Ephesians 2:8–9). This is the formal distinction from social capital models where "trust" accumulates endogenously. In our model, the deepest source of transformation comes from outside the system entirely.

2. **Common grace is persistent and universal.** $\bar{G} > 0$ for all agents, always. Even if the agent invests nothing in spiritual growth ($i^S = 0$), the steady-state soul is not zero. For an

agent who also neither cheats, it equals $\bar{G}/(\delta^S - \xi)$; for a fully negligent agent who cheats to the moral limit, corrosion lowers it to $(\bar{G} - \rho\bar{\theta})/(\delta^S - \xi - \rho\bar{\theta})$, which remains strictly positive under the calibration. This is the "grace floor": God sustains a minimal spiritual awareness, a residual conscience, even in the most negligent agent. Common grace keeps evil in check and sustains creation. Without it, the low-S equilibrium would collapse into something far worse.

3. **Effectual grace is particular and stochastic.** ε_t^G is not given to all agents in all periods. It represents God's sovereign decision to regenerate specific hearts at specific times. The timing is not predictable from within the model, but is genuinely exogenous. Christ died so that all who believe in Him would have eternal life (John 3:16), and even the faith to believe is granted by grace, not generated by human will.

4. **Non-negativity.** $\varepsilon_t^G \geq 0$: grace is always a gift, never a punishment. God may allow spiritual decay (through δ^S) but does not actively destroy the soul.

Remark (Common grace and natural law). The observable effects of common grace ($\bar{G} > 0$) closely resemble what the natural law tradition, from Aristotle through Aquinas, attributes to innate moral reason. Both predict that even unregenerate agents exhibit baseline moral behavior: a residual conscience, a sense that theft is wrong, a capacity for elementary cooperation. The economic predictions are equivalent. But the *cause* differs fundamentally. Natural law locates baseline morality in human nature itself, implying an inherent capacity that survives the Fall intact. Common grace locates it in God's ongoing sustaining act: "He causes his sun to rise on the evil and the good, and sends rain on the righteous and the unrighteous" (Matthew 5:45). The distinction matters for the model. If baseline morality is inherent, then it is structurally permanent meaning that no force can push S below the natural law floor. If it is sustained by grace, then it depends on God's continued action, which is freely given but not owed. We adopt the grace interpretation because it preserves divine sovereignty and because it is consistent with the historical observation that moral floors can erode: societies do descend into genocidal evil, which a fixed natural law floor struggles to accommodate.

Remark (Grace and technology shocks). Every economist who reads this will object: "You cannot have an exogenous variable driving your model! That is not a solution: it is an admission of ignorance!" We reply that economics already works this way: Solow (1956) and the Real Business Cycle program model technology as an exogenous stochastic process and study its consequences without first explaining where the shocks come from. Grace receives the same treatment here. But we are careful not to overstate the parallel. What licenses the RBC use of an exogenous shock is not the mere act of positing it; it is that the resulting model is held to a quantitative standard: it must match specific moments of the data. We accept that standard. Positing exogenous grace buys nothing on its own; the model earns its keep only insofar as it accounts for patterns the standard model misses. We return to this discipline, and to what those patterns are, in Section 9.

Remark (The horizon T, and why the afterlife is not a terminal payoff). The agent's problem ends at T, the close of mortal life; there is no continuation value V_{T+1} representing heaven or hell. A reader may ask whether this omits something essential, since for the Christian the afterlife is

hardly a detail. We make the choice deliberately, and for a reason internal to the model rather than mere convenience. A terminal afterlife payoff would turn devotional investment i^S into an asset with a future return (one would pray, give, and worship in order to secure the payoff), and that is the instrumentalized, works-based logic that Section 3.8 identifies as a corruption of grace. Building the afterlife in as a reward would, in the model's own terms, *lower* the soul state it was meant to raise. This is consistent with the paper's selection principle for which theological elements to formalize: we include an element when it has testable economic consequences in mortal life, is not already supplied by the standard model, and can be represented without making salvation something the agent earns. Grace meets all three tests. A heaven-and-hell payoff fails the third. This does not make the afterlife economically inert: belief in it plausibly acts as a *preference shifter* (raising the perceived return to devotion, or entering through the social externality), and Barro and McCleary (2003) find that belief in hell predicts growth. In our framework such belief would enter as a force that moves S, not as a terminal term in the objective. The theological claim the model rests on is that grace transforms life *before* death, fitting the agent for what lies beyond it; the model formalizes that transformation and leaves the eschatology to theology.

4.6 Optimality conditions

The agent chooses $\{c_t, g_t, i_t^B, i_t^M, i_t^S, a_{t+1}, \theta_t\}_{t=0}^{T}$ to maximize (1) subject to (3)–(7). All multipliers quoted in this section are in *current value*, that is, in units of period-t utility; Appendix A defines the convention precisely and uses it consistently. The key first-order conditions are:

Remark (Two versions of the problem: lifecycle and stationary). The model comes in two versions, and we keep them distinct. The *lifecycle* version is the one just stated: a finite horizon T, age-dependent body depreciation δ_t^B, and the terminal conditions $B_T = 0$ and $a_{T+1} \geq 0$. It generates the first-order and co-state equations and is the natural home of the individual's story. A finite lifecycle, however, has no time-invariant steady state: the co-states wind down as T approaches and the old save less, exactly as they should. Whenever we speak of steady states and multiple equilibria (Section 5, Proposition 7, and the simulations), we therefore use the *stationary* version: constant δ^B, an infinite horizon (equivalently, a stationary population of overlapping generations whose cross-section reproduces itself), and the same preferences and constraints otherwise. The optimality conditions are identical in form; only the terminal conditions differ. The steady states of the paper are steady states of the stationary version, and no result asks the lifecycle version to sit still.

Consumption Euler equation. Combining the consumption first-order conditions at t and $t+1$ with the first-order condition for the financial asset a_{t+1}:

$$c_t^{-\sigma(S_t)} = \beta(S_{t+1})\,(1 + r_{t+1})\,c_{t+1}^{-\sigma(S_{t+1})} \tag{9}$$

Both the discount factor and the curvature depend on S, so the consumption path is jointly determined with the spiritual trajectory. The asset is what makes this equation a theorem rather than an assertion: it is the margin along which consumption today trades against consumption tomorrow.

Giving ratio:

$$\frac{g_t}{c_t} = \psi(S_t)^{1/\sigma(S_t)} \tag{10}$$

When $S = 0$, the agent gives nothing. As S rises, generosity emerges endogenously.

Honesty. The corruption choice is bounded on both sides, $0 \leq \theta_t \leq \bar{\theta}(1 - S_t)$, so the first-order condition carries a Kuhn–Tucker multiplier for each bound:

$$\lambda_t \Pi_t - \mu_t^S \rho - \nu_t + \kappa_t = 0, \qquad \kappa_t \theta_t = 0, \qquad \nu_t \left[\bar{\theta}(1 - S_t) - \theta_t\right] = 0, \tag{11}$$

with $\nu_t \geq 0$ on the moral ceiling and $\kappa_t \geq 0$ on the zero floor. The term $\lambda_t \Pi_t$ is the marginal benefit of cheating: one more unit of θ raises income by Π_t, valued at shadow price λ_t. The term $\mu_t^S \rho$ is the *spiritual cost*, the value of the soul that corrosion destroys. Because both benefit and cost are linear in θ_t, the choice is a comparison between them. If the spiritual cost exceeds the benefit, $\mu_t^S \rho > \lambda_t \Pi_t$, the agent sets $\theta_t = 0$ with $\nu_t = 0$ and $\kappa_t = \mu_t^S \rho - \lambda_t \Pi_t > 0$: the moral ceiling is slack and the agent *chooses honesty even though dishonesty is permitted*, because the future spiritual loss outweighs the immediate gain. (The floor multiplier κ_t is what lets the first-order condition hold at this corner; without it the condition could not be satisfied there at all.) If $\mu_t^S \rho < \lambda_t \Pi_t$, the ceiling binds: $\theta_t = \bar{\theta}(1 - S_t)$, $\kappa_t = 0$, and $\nu_t = \lambda_t \Pi_t - \mu_t^S \rho > 0$. Exact indifference, $\mu_t^S \rho = \lambda_t \Pi_t$, is the knife-edge in between. The chosen-honesty case is precisely what the always-binding formulation lacked, and it is what gives the patience channel of Proposition 1 its force: a patient agent (high β) and an agent whose soul is valuable (high μ_t^S) both weigh the corrosion cost more heavily and cheat less.

Equal marginal returns to investment:

$$\mu_t^B \phi^{B\prime}(i_t^B) = \mu_t^M \frac{\partial \phi^M}{\partial i_t^M} = \mu_t^S \phi^{S\prime}(i_t^S) = \lambda_t \tag{12}$$

The agent allocates investment across body, mind, and soul until the marginal returns are equalized. This is a standard optimality condition. The novelty is that spiritual investment competes with (and complements) physical and human capital investment.

The shadow value of the soul:

$$\mu_t^S = \underbrace{\beta'(S_{t+1})\, V_{t+1}}_{\text{patience}} + \beta(S_{t+1}) \Big[\underbrace{\frac{\partial u_{t+1}}{\partial \sigma} \frac{\partial \sigma}{\partial S}}_{\text{satiation}} + \underbrace{\frac{\partial u_{t+1}}{\partial \psi} \frac{\partial \psi}{\partial S}}_{\text{generosity}} - \underbrace{\nu_{t+1} \bar{\theta}}_{\text{honesty}} + \mu_{t+1}^S (1 - \delta^S) \Big] \tag{13}$$

where V_{t+1} is the continuation value from $t+1$ onward. This is the central equation of the model. The soul is a *meta-variable* that transforms preferences, time horizons, social orientation, and moral constraints simultaneously, and the four braced terms are the four channels. Their signs deserve comment. The patience term is positive: a higher soul raises the discount factor applied to the *entire* future, scaling up everything the agent still has to live for. The generosity term is positive. The satiation term is a level effect whose sign depends on the units of consumption (Remark 4.1). The honesty term enters *negatively* when the moral ceiling binds: at the margin, a higher soul tightens conscience and withholds rents the agent still wanted (no bookkeeping can make a binding constraint privately valuable). Conscience is privately costly exactly when

temptation is still alive, and socially productive all the while (Proposition 4): the wedge between the two is why no market supplies it. The term also vanishes for a high-S agent, who no longer wishes to cheat ($\nu = 0$), and drags hardest at low S, so the private return to the soul is weakest where society is trapped. That asymmetry deepens the low-soul trap of Proposition 7.

5 Four Channels and Seven Propositions

The soul transforms economic behavior through four channels. Each channel generates propositions, and one distinction should be fixed before the first of them. A *proposition*, here as anywhere in theory, is a conditional statement: given the stated assumptions and equilibrium conditions, the formal result follows. It is correctly or incorrectly derived, nothing else. Its *empirical* counterpart is a hypothesis about observables: proxies for the soul state (behavioral measures, not pew counts; Section 9.2) should relate to observable outcomes in the direction the proposition dictates, and that hypothesis is supported or rejected by data. We state each proposition with its conditions, prove it under them, and leave the reader in no doubt which claims are theorems of the model and which are wagers about the world.

5.1 Channel 1: Patience

The soul transforms the time horizon. A person grounded in Christ thinks long-term; e.g., stewardship, investment, building for the next generation. An agent driven by desire discounts the future heavily and grabs what is available now.

Formally, $\beta(S) = \underline{\beta} + (\bar{\beta} - \underline{\beta})S$ rises with S: when S is high, β is close to 1 and the agent is patient. When S is low, β is low and the agent is impatient. The mechanism has a mainstream precedent: Becker and Mulligan (1997) show that agents can invest resources to raise their discount factor, producing divergent steady states in wealth. Our contribution is to identify the *source* of that investment (devotional practice) and the *shock* that catalyzes it (grace).

Proposition 1 (The Patience Effect)**.** *Suppose the corruption choice satisfies* $0 \le \theta_t \le \bar{\theta}(1-S_t)$*, the discount factor satisfies* $\beta'(S) > 0$*, and the ratio of the spiritual cost of cheating to its marginal benefit,* $\mu^S \rho / \lambda\Pi$*, is nondecreasing in* S*. Then the optimal corruption policy is weakly decreasing in the soul state,*

$$\frac{\partial \theta^*(S)}{\partial S} \le 0,$$

and takes a threshold form: there exists S^θ *such that* $\theta^*(S) = \bar{\theta}(1-S)$ *for* $S < S^\theta$ *and* $\theta^*(S) = 0$ *for* $S > S^\theta$*. The decline is strict wherever* $\theta^*(S) > 0$*.*

Proof sketch. The objective is linear in θ_t (both the rent gain $\lambda_t \Pi_t$ and the corrosion cost $\mu_t^S \rho$ scale one-for-one with θ), so by the Kuhn–Tucker conditions (11) the choice is at a corner except at exact indifference: cheat to the ceiling when $\lambda_t \Pi_t > \mu_t^S \rho$, abstain entirely when the inequality reverses. The threshold form then follows from the monotonicity assumption on the cost-benefit ratio, which is what the patience channel supplies: the spiritual cost is the discounted future value of the soul that dishonesty corrodes, and since μ_t^S carries the factor $\beta(S_{t+1})$ (equation (13)), it rises with the discount factor. An impatient agent (low S, low β) discounts the corrosion away and appropriates rents for immediate gain; a patient agent finds the corrosion cost looms large

and declines to cheat even when the moral constraint would permit more. Below the threshold the binding ceiling gives $\partial\theta^*/\partial S = -\bar{\theta} < 0$; above it the policy sits at zero and the derivative is zero, which is why the global statement is weak rather than strict. Full proof in Appendix B. □

Remark (The tenure of corruption). A dynamic corollary, stated informally because it needs assumptions beyond the comparative static: if the soul drifts upward under persistent grace ($\bar{G} > 0$) and the environment is otherwise stationary, an agent currently cheating exits corruption when S first crosses S^θ, and the expected remaining duration of the corrupt spell is decreasing in the current soul state. Making this a theorem requires specifying the stochastic process for S and for the rent pool Π_t; the derivation companion states the required conditions. The empirical implication is a corruption-exit hazard that rises with proxies for S.

5.2 Channel 2: Satiation

The soul transforms how much we need. A person grounded in Christ is more easily satisfied with material goods; "godliness with contentment is great gain" (1 Timothy 6:6). An agent driven by desire is never satisfied.

Formally, $\sigma(S) = \underline{\sigma} + (\bar{\sigma} - \underline{\sigma})S$ rises with S: higher curvature means the marginal utility of consumption diminishes quickly and the agent saves more. The link between moral identity and consumption restraint is documented by Bénabou and Tirole (2011), who show that agents with stronger moral self-image resist temptation more effectively, investing in identity as an "asset" that disciplines future behavior.

Proposition 2 (The Satiation Effect)**.** *Suppose $\beta'(S) > 0$, the satiation condition $\partial^2 u/\partial c\,\partial S \leq 0$ holds on the relevant domain (Remark 4.1), and current resources and expected returns are held fixed. Then a higher soul state raises the total transfer of resources to the future, financial saving plus investment,*

$$\frac{\partial}{\partial S^*}\left(a^* + i^{B*} + i^{M*} + i^{S*}\right) \geq 0,$$

with strict inequality when the patience effect dominates the income and substitution effects generated by the other preference channels. Under the baseline calibration it does.

Proof sketch. With the financial asset in the budget constraint (3), saving is defined and the Euler equation (9) disciplines its path. A higher $\beta(S)$ tilts the consumption profile toward the future: patience alone raises the resources carried forward. The satiation condition adds that a higher soul state weakly lowers the marginal utility of present consumption, reinforcing the tilt rather than fighting it. What keeps the general statement conditional is the generosity channel: rising $\psi(S)$ pulls resources into giving, which competes with saving, and a rising σ moves income and substitution effects in opposite directions when returns vary. The dominance condition says the patience effect wins that contest; it is an assumption, verified numerically under the baseline calibration rather than derived, and we flag it as such. Details in Appendix A. □

5.3 Channel 3: Generosity

The soul transforms how we treat others. A Christ-centered person gives voluntarily to the poor, to the neighbor, to the stranger. This is not redistribution imposed by the state. It is redistribution that flows naturally from a transformed heart.

Formally, $\psi(S) = \psi_0 S^\eta$ rises with S. The giving-to-consumption ratio $g/c = \psi(S)^{1/\sigma(S)}$ increases monotonically. Andreoni (1990) formalized this mechanism as "warm-glow giving": agents derive intrinsic utility from the act of giving, independent of the public good produced. Our framework endogenizes Andreoni's warm glow. It is not a fixed taste parameter but a function of the soul state.

Warm glow and the hidden hand. Modeling generosity as a utility term invites an obvious objection. Scripture tells the giver not to let the left hand know what the right hand is doing (Matthew 6:3), yet our agent gives *because* giving yields utility. Have we made charity into self-interest? The answer turns on which giving $\psi(S)\,u(g)$ represents. Andreoni's warm glow is a *private, internal* return, not a public one; it is precisely not the reward the Pharisee seeks when he gives to be seen (Matthew 6:2). A reputational or self-signaling motive would put the *audience* in the utility function, and our model deliberately does not. What remains is the gladness of a cheerful giver (2 Corinthians 9:7), which Scripture commends rather than condemns. The representation is positive, not normative: the transformed agent gives *as if* giving entered her utility because giving has in fact become her desire, not because she calculates a glow. And the model already refuses the works-righteousness shortcut, since giving does not build the soul (there is no g term in equation (7)); the only return to charity is the contemporaneous warm glow, never a spiritual yield. We concede the limit. Pure agape, giving with no return to the self of any kind, is something no utility representation can fully hold, and we do not claim $\psi(S)\,u(g)$ exhausts it. It captures the economic shadow of charity, the part that moves resources and that we can measure, and leaves the rest to grace.

Proposition 3 (Generosity Emergence)**.** *Under the stated monotonicity conditions ($\psi'(S) > 0$, $\sigma'(S) \geq 0$, $\psi_0 \leq 1$) and an interior allocation, the optimal ratio of voluntary giving to own consumption is zero at $S = 0$ and strictly increasing in the soul state:*

$$\frac{\partial}{\partial S}\left(\frac{g^*}{c^*}\right) > 0.$$

Proof sketch. From equation (10), $g/c = \psi(S)^{1/\sigma(S)}$; differentiating $\ln[g/c]$ with respect to S shows the ratio rises globally whenever $\psi \leq 1$, which $\psi_0 \leq 1$ guarantees on $[0, 1]$ (Appendix B). The proposition is deliberately stated in the giving *share*, because that is what the first-order conditions pin down; the *level* of giving also depends on how consumption itself moves. Under the baseline calibration the share effect dominates the fall in consumption over the relevant range, so the level of g^* rises too, but that is a calibration fact, not a theorem. □

Corollary 1 (When forced redistribution is redundant)**.** *Let T^{SP} denote the transfer a social planner would mandate. If the soul state is high enough that voluntary giving covers it, $g^*(S) \geq T^{SP}$, forced redistribution is redundant: the transfers the planner wants already flow, without the apparatus of compulsion. Under the baseline calibration, giving rises from roughly 5 percent of consumption at the low equilibrium to roughly 25 percent at $S = 0.70$, so the condition is met for any plausible T^{SP} well before S reaches one.*

This proposition has radical policy implications. Government welfare programs, affirmative action schemes, and forced redistribution are all substitutes for generosity in a low-soul society.

They are necessary precisely because voluntary giving is insufficient. But they are also inefficient, corruptible, and dignity-reducing. A high-soul society generates the same transfers (and more) through voluntary action, without the bureaucratic overhead or the resentment.

5.4 Channel 4: Honesty

The soul imposes an internal moral constraint. A Christ-centered person *cannot* cheat, not because the law prevents it, but because an internal conviction makes it impossible.

Formally, $\theta_t \leq \bar{\theta}(1 - S_t)$. When $S = 1$, the agent cannot cheat at all. When $S = 0$, the constraint is slack. Mazar et al. (2008) provide direct experimental evidence for this internal ceiling: subjects cheat only up to the point where dishonesty would force a negative revision of their moral self-concept. Our soul state S is the variable that determines where that ceiling sits.

Proposition 4 (The Honesty Dividend)**.** *Suppose corruption diverts each agent's effective labor in proportion θ_i, all honesty ceilings bind, and θ_i is uncorrelated with $M_i B_i$ across agents. Then measured TFP is $A\,[1 - \bar{\theta}(1 - \bar{S})]^{\alpha}$, strictly increasing in the average soul state, with exact semi-elasticity*

$$\frac{\partial \ln TFP}{\partial \bar{S}} = \frac{\alpha\bar{\theta}}{1 - \bar{\theta}(1 - \bar{S})} > 0.$$

Proof sketch. While corruption benefits the individual (equation 3), it diverts resources from productive use at the aggregate level. A corrupt official who appropriates infrastructure funds raises her own consumption but destroys the public good. Aggregate effective labor is therefore $\bar{N} = \int M_i B_i (1 - \theta_i)\, di$, and aggregate output is $Y = A\bar{N}^{\alpha} K^{1-\alpha}$. Measured TFP (the Solow residual, computed from observed inputs $N = \int M_i B_i\, di$) equals $A(\bar{N}/N)^{\alpha}$. Two assumptions turn the heterogeneous ratio into an average: no correlation between θ_i and effective labor $M_i B_i$ (otherwise the covariance enters), and binding ceilings $\theta_i = \bar{\theta}(1 - S_i)$, whose linearity in S_i lets individual soul states aggregate exactly to $\bar{S}$. Then $\bar{N}/N = 1 - \bar{\theta}(1 - \bar{S})$ and the exact formula follows by differentiation. When the corruption wedge $\bar{\theta}(1 - \bar{S})$ is small, the semi-elasticity is approximately

$$\Delta \text{TFP} \approx \alpha\bar{\theta}\, \Delta\bar{S} \tag{14}$$

With $\alpha = 0.65$ and $\bar{\theta} = 0.3$, this local approximation gives $\alpha\bar{\theta} = 0.195$: a one-unit increase in the average soul state raises the measured TFP level by roughly 20%. The exact formula gives more, about 26% over the full unit interval, because the denominator shrinks as honesty recovers; the simulations use the exact expression. Either way it is a *level* effect, not a permanent change in the growth *rate*; like a one-off technology improvement, it shifts the level of the Solow residual. □

5.5 Additional propositions

Proposition 5 (The Forgiveness Equilibrium)**.** *Consider a symmetric two-group game with strategies Forgive (F) and Revenge (R) and payoffs $\pi^{FF}(S) = \beta(S)V^{peace} + \psi(S)\Omega$, $\pi^{RF}(S) = R_0 - \psi(S)\Gamma$, $\pi^{FR}(S) = -L + \beta(S)V^{recovery}$, and $\pi^{RR} < 0$. Define the gain to forgiving against a forgiver, $\Delta_F(S) = \pi^{FF} - \pi^{RF}$, and against a revenger, $\Delta_R(S) = \pi^{FR} - \pi^{RR}$; both are strictly increasing in S. Let S^F solve $\Delta_F = 0$ and S^R solve $\Delta_R = 0$ where these are interior. Then:*

(i) mutual forgiveness (F, F) is a Nash equilibrium if and only if $S \geq S^F$, and forgiveness is strictly dominant, making (F, F) the unique equilibrium, for $S > \max\{S^F, S^R\}$; (ii) mutual revenge (R, R) is a Nash equilibrium if and only if $S \leq S^R$, and revenge is strictly dominant for $S < \min\{S^F, S^R\}$; (iii) between the thresholds the game is a coordination game if $S^F < S^R$ (both (F, F) and (R, R) are equilibria, and history selects) and a hawk–dove game if $S^R < S^F$ (only the asymmetric profiles are equilibria: one group revenges while the other's forbearance is exploited).

Proof sketch. Δ_F rises in S because patience values peace and generosity both rewards forgiving (Ω) and recoils from harming (Γ); Δ_R rises because patience values recovery. Monotonicity plus the intermediate value theorem give unique thresholds, and the regime classification is then just the definition of Nash equilibrium applied bound by bound; one payoff comparison alone cannot establish dominance, which requires both. Under the baseline calibration $S^F \approx 0.28$ while S^R lies below zero (recovery after victimization, $-L + \beta V^{\text{recovery}}$, beats mutual destruction π^{RR} at every S), so the operative case is (iii) hawk–dove: below S^F mutual forgiveness cannot be sustained, and every equilibrium has at least one group taking revenge while the forgiving group absorbs the blow. Revenge is never literally dominant; what fails below the threshold is *mutual* forgiveness. Deeper wounds (larger L relative to the ruin of mutual destruction) push S^R above zero and open the coordination region, where revenge becomes self-fulfilling: each group revenges because it expects the other to. Both regimes connect this game to Proposition 7: in the intermediate range the outcome is decided not by payoffs alone but by which expectations a society coordinates on, and that is exactly what a critical mass of transformed agents changes. Full proof in Appendix B. □

Proposition 6 (The Unemployment–Soul Nexus)**.** *In a wage- and price-setting model of structural unemployment, soul-dependent markups reduce structural unemployment. Formally, $\partial \bar{u} / \partial \bar{S} < 0$.*

Setup. We embed the soul in a standard model of structural unemployment with two bargaining parties. A continuum of monopolistically competitive firms set prices above marginal cost, generating an aggregate price markup μ^P. Wages are set by bargaining between a representative firm and a union over the joint surplus from a match. The bargained wage exceeds the competitive level by a wage markup μ^W; in the Nash bargaining solution this markup is increasing in the union's bargaining power $\beta_{\text{union}} \in (0, 1)$, the weight the union carries in the Nash product, with threat points given by the firm's outside option (lost output) and the union's (its members' fallback income). With a labor-tax wedge τ, the wage-setting and price-setting schedules intersect at a structural (involuntary) unemployment rate

$$\bar{u} = 1 - \frac{1 - \tau}{(1 + \mu^P)(1 + \mu^W)},$$

the familiar result that unemployment rises in both markups and in the tax wedge. The two parties bargain over the wage; firms additionally choose the price markup. Both decisions are made by agents with a soul state.

Proof sketch. The soul enters through both parties. A union whose members have higher S internalizes, through the generosity channel $\psi(S)$, the cost that aggressive wage demands impose

on the unemployed outsiders, and so moderates its demand: $\mu^W(S) = \frac{\beta_{\text{union}}}{1-\beta_{\text{union}}}(1-\psi(S)\kappa)$, with $\kappa > 0$ scaling how strongly other-regard tempers the wage claim. A firm whose owners and managers have higher S extracts a smaller price markup over cost: $\mu^P(S) = \mu_0^P(1 - S\zeta)$, with $\zeta > 0$. These linear forms need explicit admissibility restrictions, $\psi_0\kappa \leq 1$ and $\zeta \leq 1$, so that both markups remain nonnegative over the whole range of S; the baseline calibration satisfies them comfortably ($\psi_0\kappa = 0.4$, $\zeta = 0.3$). Since $\bar{u}$ is increasing in both μ^P and μ^W, and both are decreasing in the (average) soul state, $\partial\bar{u}/\partial\bar{S} < 0$. A fuller treatment of the bargaining problem, including the derivation of μ^W from the Nash product, is developed in the companion appendices of the broader project; here we record the reduced-form result that connects the soul to structural unemployment. □

Proposition 7 (Critical Mass and Multiple Equilibria)**.** *Consider the symmetric steady-state condition* $(\delta^S - \xi)\bar{S} = R(\bar{S})$*, where* $R(\bar{S}) = \phi^S(i^{S*}(\bar{S})) + \bar{G} - \rho\,\theta^*(\bar{S})$ *is the replenishment curve and* $\delta^S - \xi > 0$*. Assume* R *is continuously differentiable, nondecreasing, and* S-shaped*:* $R'(\bar{S}) < \delta^S - \xi$ *at low* $\bar{S}$*,* $R'(\bar{S}) > \delta^S - \xi$ *on an intermediate interval,* $R'(\bar{S}) < \delta^S - \xi$ *again at high* $\bar{S}$*, with endpoint conditions* $R(0) > 0$ *and* $R(1) < \delta^S - \xi$*. Then the steady-state equation has exactly three solutions* $\bar{S}_L < \bar{S}^* < \bar{S}_H$*: the outer two are locally stable and the middle one is unstable. A society starting below* $\bar{S}^*$ *converges to the low-soul equilibrium (high corruption, low trust, low growth); above it, to the high-soul equilibrium.*

Proof sketch. The endpoint conditions place R above the depreciation line at zero and below it at one; the assumed slope pattern then forces exactly three crossings. The shape is doing real work and we state it as an assumption rather than pretend to derive it: concavity of ϕ^S alone would deliver at most two crossings, and it is the full sigmoid, flat then steep then flat, that produces three. The economics behind the assumption is the investment complementarity: at low $\bar{S}$ the soul's shadow value is small and $i^{S*} \approx 0$, so replenishment is flat; at intermediate $\bar{S}$ the social multiplier and the rising value of the soul engage, and the receding corrosion term $-\rho\theta^*(\bar{S})$ adds to the slope as cheating recedes; at high $\bar{S}$ the concavity of ϕ^S imposes diminishing returns and $\theta^* \to 0$ leaves nothing more to recover. The calibrated policy function satisfies the assumption (Figure 2), but we have not derived it from primitives, and a reader who wants Proposition 7 to rest on weaker ground should read the S-shape as the price of the result. Stability is checked against the discrete dynamics $\bar{S}_{t+1} = F(\bar{S}_t)$ with $F(\bar{S}) = \bar{S} + R(\bar{S}) - (\delta^S - \xi)\bar{S}$: local stability requires $|F'| < 1$, that is, $(\delta^S - \xi) - 2 < R' < \delta^S - \xi$. The upper inequality is the familiar slope comparison $L' > R'$; the lower one holds automatically because R is nondecreasing and $\delta^S - \xi < 2$, so for monotone replenishment the slope comparison is the whole story. It holds at $\bar{S}_L$ and $\bar{S}_H$ and fails at $\bar{S}^*$. Full construction in Appendix B. □

Remark (A literal critical mass). The proposition's threshold lives in the *average* soul state $\bar{S}$; turning it into a count of people requires heterogeneity. The mean-field arithmetic is immediate: if a fraction π of agents is transformed to S^H while the rest remain at S^L, the average crosses the tipping point when $\pi > \pi^* = (\bar{S}^* - S^L)/(S^H - S^L)$. Under the baseline calibration ($\bar{S}^* = 0.65$, $S^L = 0.20$, $S^H = 0.85$) that is roughly two-thirds of the population, a majority, not a remnant. The figure is illustrative rather than exact (it holds the untransformed fixed, whereas in the model the social multiplier and common grace lift them as $\bar{S}$ rises, easing the requirement), but the qualitative point stands and is sobering: tipping a society is not the work of a devout sliver.

6 Two Cities, Two Equilibria

Augustine's *De Civitate Dei* was written between 413 and 426 AD, as Rome was sacked by the Visigoths and pagans blamed Christianity for the Empire's decline. Augustine's response was not to defend Rome but to reframe the question. Rome, he argued, was a city built on the love of self, or the *libido dominandi*, the lust for domination. Its decline was not caused by Christianity but by the same disordered desires that built it. The City of God, by contrast, is built on the love of God and of neighbor for God's sake. It exists alongside the City of Man in every generation, often invisible, but it is the only city that endures (Augustine, 426).

The parallels with our model are exact:

	City of Man ($\bar{S}_L$)	**City of God** ($\bar{S}_H$)
Driving force	Self-love (*amor sui*)	Love of God (*amor Dei*)
Desire	Unbounded, insatiable	Bounded by contentment
Time horizon	Short-term, impatient	Long-term, patient
Social orientation	Self-interest, exploitation	Generosity, service
Moral constraint	External only (law, punishment)	Internal (conscience, conviction)
Model parameter	Low S, low β, low σ, $\psi \approx 0$	High S, high β, high σ, high ψ
Corruption	High ($\theta \approx \bar{\theta}$)	Low ($\theta \approx 0$)
Savings	Low	High
Forgiveness	Unsustainable; revenge in equilibrium	Forgiveness dominant
Growth	Stagnant	Flourishing

Augustine understood what our model formalizes: that the difference between a flourishing society and a decaying one is not primarily institutional but *dispositional*. It depends on what the citizens love. "Two cities have been formed by two loves: the earthly by the love of self, even to the contempt of God; the heavenly by the love of God, even to the contempt of self" (*De Civitate Dei*, XIV.28).

Kierkegaard sharpens the diagnosis. In *The Sickness Unto Death*, he defines despair as the condition of a self that "does not will to be itself" or, equivalently, that "wills not to be the self that it truly is" (Kierkegaard, 1849). For Kierkegaard, the self is constituted by a Power outside itself: God. Despair is the rejection of that relationship: the self tries to construct itself on its own terms, becoming a fabrication, a mere image that is not the person God intended. In the language of our model, this is precisely the low-S agent. When $S \to 0$, the agent rejects the divine constitution of the self and builds an identity around desire (power, status, wealth). The result is not freedom but bondage: insatiable consumption (σ low), impatience (β low), dishonesty (θ high), and indifference to others ($\psi \approx 0$). The agent becomes less than human, not more. And when an entire society is composed of such agents, it is easy to see why it turns to evil, that ultimately destroys itself. The low-S equilibrium is in Kierkegaard's sense, a society in despair: a collective of selves that have rejected the Power that constituted them, and in doing so have lost the capacity to flourish.

6.1 The critical mass and the role of the Church

Proposition 7 shows that the transition between equilibria requires a critical mass. In our calibration the tipping point in the average soul state is $\bar{S}^* \approx 0.65$, and translating the average into people (the mean-field arithmetic in the remark following Proposition 7) puts the required share of transformed agents at roughly two-thirds of the population before the social multiplier pulls the entire society upward.

This has implications for the Church. The Church is the institution through which grace is proclaimed, fellowship sustained, and devotional investment (i^S) made most productive. The social multiplier ξ is highest within church communities where weekly worship, small groups, accountability structures, shared Scripture all amplify the individual's spiritual growth. The model predicts that a society's trajectory depends critically on the size, vitality, and theological seriousness of its churches.

6.2 Historical evidence: Rwanda and South Africa

Rwanda is suggestive, and we present it with care. The country was overwhelmingly Christian *before* 1994, and its churches were widely implicated in the genocide rather than innocent of it (Longman, 2010). That is itself a datum for our framework: nominal affiliation is not a high soul state (Section 3.7), and the most-churched nation in Africa was not thereby spared. What changed afterward is contested, and we do not overstate it. The Gacaca courts, some 12,000 community tribunals that processed over a million cases, were a state-driven mechanism, and Rwanda's recovery is bound up with an authoritarian order that has enforced national unity from above. But alongside the coercion there was a documented turn in many churches toward confession, reconciliation, and forgiveness, and the economy has grown at 7 to 8 percent a year since 2000. We read Rwanda not as proof but as a case consistent with the model's prediction that a society can move toward the forgiveness equilibrium when transformation reaches past the courtroom into the heart.

South Africa is the instructive contrast. The Truth and Reconciliation Commission was *explicitly* theological in its framing, grounded by Archbishop Desmond Tutu in Christian forgiveness and the ethic of *ubuntu* (Tutu, 1999), achieved something real: it surfaced the truth and forestalled the retributive bloodshed many had feared. Our claim is therefore not that the TRC ignored the soul. It is that the transformation it modeled at the top did not propagate to the critical mass the model requires. Aggregate $\bar{S}$ stayed below the forgiveness threshold S^F, and the redistributive politics that followed ran increasingly on the low-soul logic of transfer and grievance rather than reconciliation. Two serious attempts at national healing, two different depths of transformation, two trajectories. We develop both cases, with the supporting data, in the companion book; here they serve to illustrate the mechanism, not to test it.

7 Numerical Simulations

This section presents calibrated simulations illustrating the model's quantitative predictions. The baseline calibration is summarized in Table 3.

Table 3: Baseline calibration parameters

Parameter	Interpretation	Value	Source/Justification
$\underline{\beta}$	Min patience (low soul)	0.92	Standard macro
$\bar{\beta}$	Max patience (high soul)	0.98	Standard macro
$\underline{\sigma}$	Min curvature	1.2	Standard CRRA
$\bar{\sigma}$	Max curvature	3.0	Higher = more easily satisfied
ψ_0	Scale of giving weight	0.05	Matches tithing-scale giving
η	Convexity of giving in S	1.0	Linear response
$\bar{\theta}$	Max feasible cheating	0.30	Corruption surveys
δ^S	Spiritual decay rate	0.30	Moderate decay from neglect
ρ	Soul corrosion from cheating	0.10	Modest; sin hardens the heart
$\bar{G}$	Persistent grace	0.05	Small but positive
ξ	Social multiplier	0.10	Peer effects literature
α	Labor share	0.65	National accounts

7.1 Comparative statics of the soul state

Figure 1 shows how steady-state outcomes change as S varies from near zero to near one.

Panel (c) is perhaps the most striking. At $S = 0$, the agent consumes virtually everything. As S rises the investment share climbs steeply, to more than a quarter of income by $S = 0.8$, while consumption contracts and voluntary giving emerges. This is the micro-foundation for the claim that spiritual transformation drives capital accumulation. The soul does not compete with growth, but rather *enables* it.

7.2 The two-equilibria phase diagram

Figure 2 plots the steady-state condition from Proposition 7. The straight line $L(\bar{S})$ represents net depreciation ($(\delta^S - \xi)\bar{S}$, linear). The replenishment curve $R(\bar{S})$ represents total replenishment (grace, investment response, and the soul recovered as corrosion recedes). The replenishment curve is S-shaped: flat at low $\bar{S}$ (nobody invests in soul when the social environment is hostile), then steep (complementarity kicks in), then flattening (diminishing returns). The two curves cross three times.

7.3 The welfare landscape

Figure 3 maps steady-state welfare as a function of grace ($\bar{G}$) and devotional investment (i^S).

The key insight: grace without devotion (moving right along the bottom edge) raises welfare only modestly. Devotion without grace (moving up along the left edge) is even less effective. The interaction is the key: grace opens the heart, devotion builds the habit, and together they transform the agent. This is the economic content of "work out your salvation with fear and trembling, for it is God who works in you" (Philippians 2:12–13).

7.4 GDP decomposition by soul channel

Figure 4 decomposes the GDP gain from higher $\bar{S}$ into three channels: the honesty dividend (green), human capital from patience (blue), and physical capital from savings (gold).

Figure 1: Comparative Statics of the Soul State

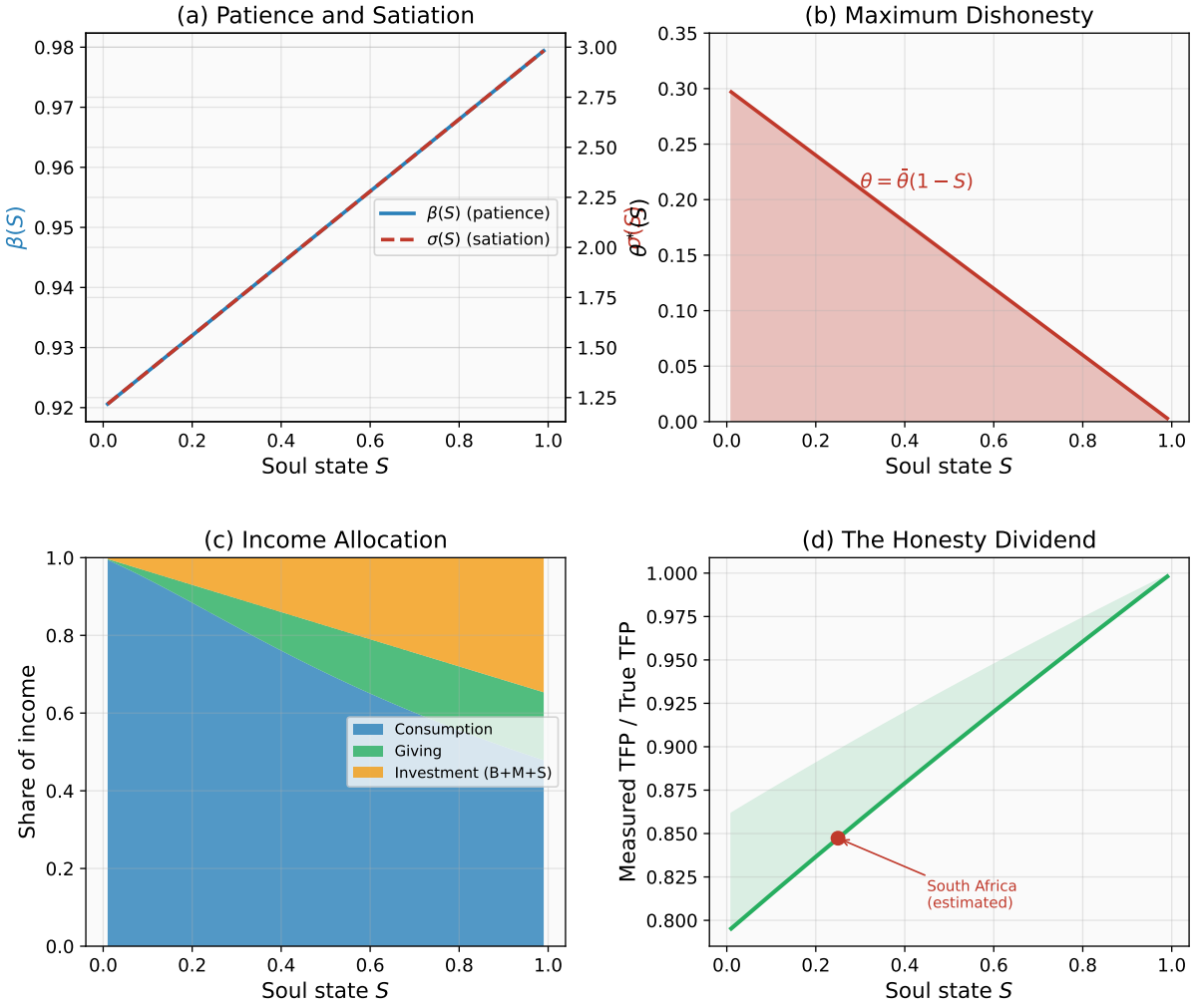


Figure 1: Comparative statics of the soul state. Panel (a): patience $\beta(S)$ and satiation $\sigma(S)$ rise with S. Panel (b): maximum dishonesty $\theta^*(S) = \bar{\theta}(1 - S)$ falls. Panel (c): the allocation of income shifts from consumption toward giving and investment as S rises. Panel (d): measured TFP rises as honesty improves (the honesty dividend).

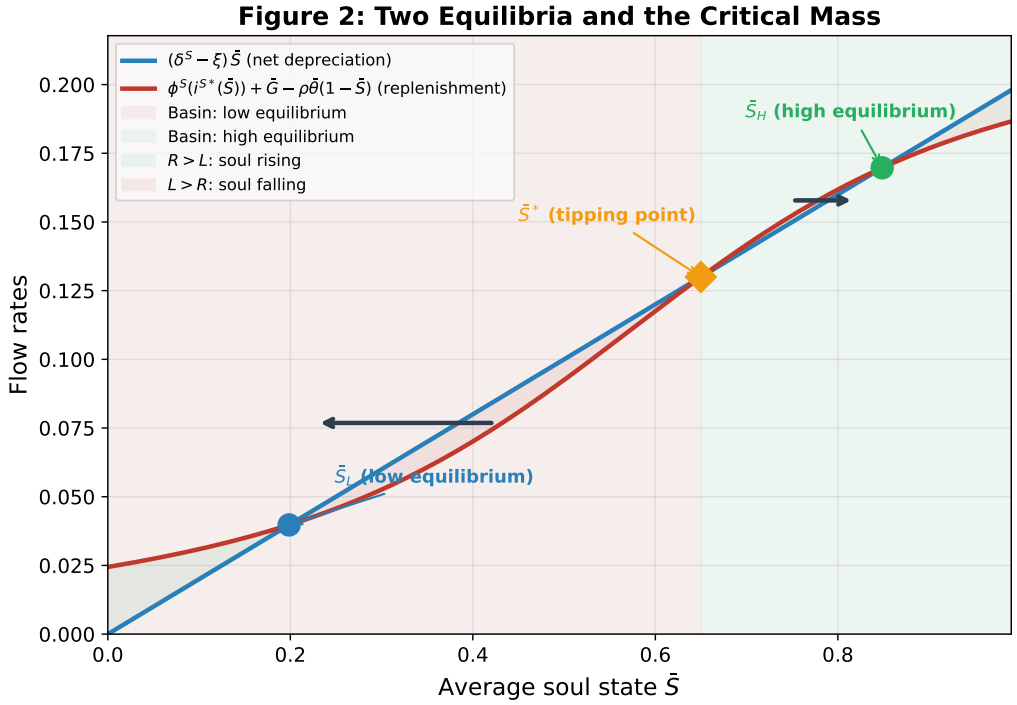


Figure 2: Phase diagram for the average soul state $\bar{S}$. The linear $L(\bar{S})$ and the S-shaped $R(\bar{S})$ cross three times: two stable equilibria, a low-soul trap ($\bar{S}_L \approx 0.20$) and a high-soul flourishing state ($\bar{S}_H \approx 0.85$), separated by an unstable tipping point, the critical mass $\bar{S}^* \approx 0.65$, which divides the basins of attraction.

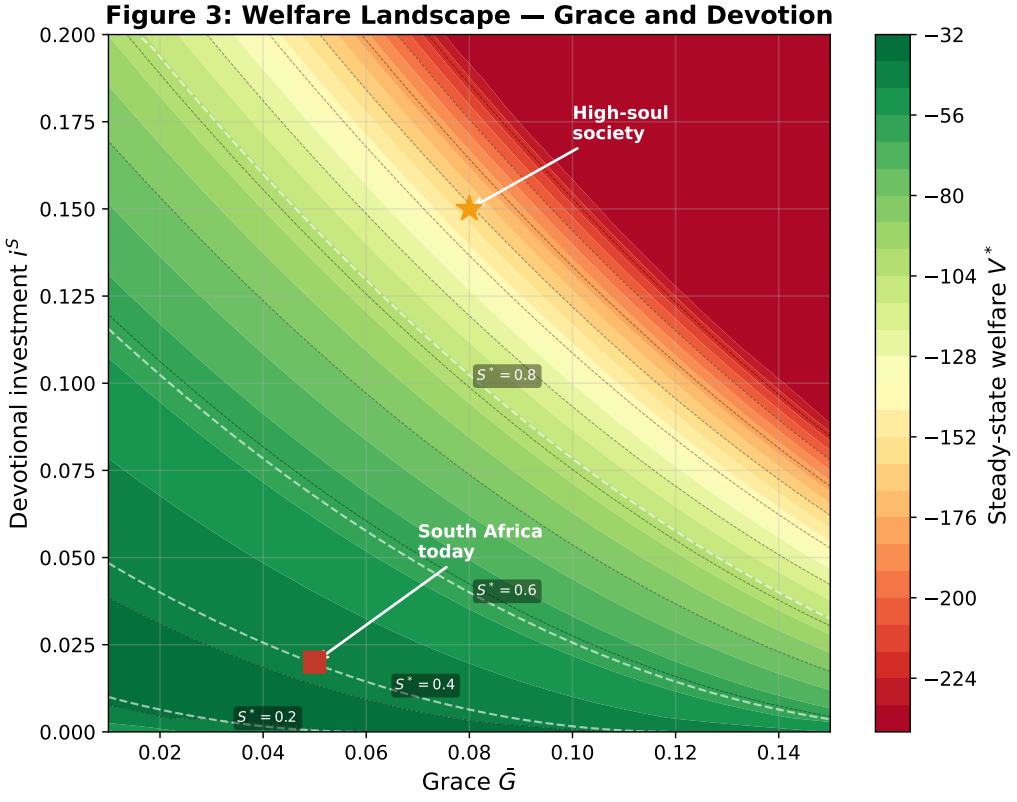


Figure 3: Welfare landscape over grace and devotional investment. Grace and devotion are complements, not substitutes. The welfare gap between a low-soul and high-soul society spans orders of magnitude.

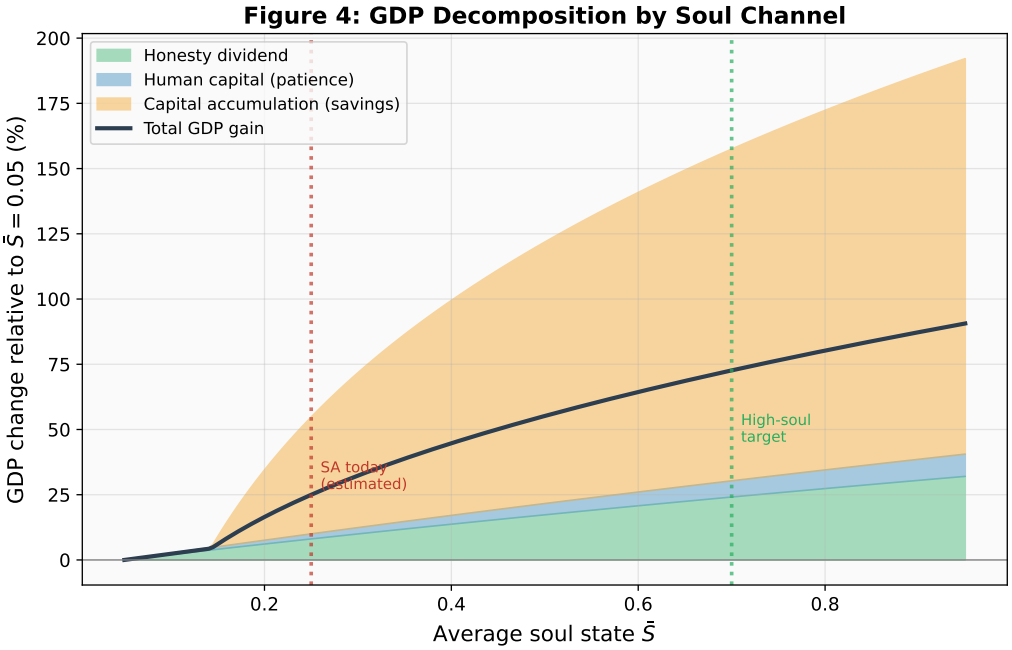


Figure 4: GDP decomposition by soul channel, relative to a baseline of $\bar{S} = 0.05$. Moving from the low-soul equilibrium ($\bar{S} \approx 0.20$) to a high-soul state ($\bar{S} = 0.70$) raises GDP by roughly 75% in this decomposition; from South Africa's estimated current level ($\bar{S} \approx 0.25$) the gain is about 60%.

The composition shifts with S. At low S, the honesty dividend dominates; eliminating corruption is the single biggest gain. At moderate S, human capital takes over as patient agents invest in education. At high S, capital accumulation compounds where high savings rates build the capital stock.

7.5 The forgiveness game

Figure 5 plots the payoff functions from Proposition 5.

The calibration reveals a sobering pattern. At $S = 0$, revenge against a forgiving opponent is overwhelmingly attractive, so whoever forgives is exploited and mutual forgiveness collapses. This is the world of vendetta and "our turn to eat." But above S^F, the gap widens rapidly: at $S = 0.8$, revenge is not just inferior, it is unthinkable for a high-S agent. Forgiveness is not weakness. It is the optimal strategy of the transformed soul.

7.6 Transition dynamics

Figure 6 simulates a society starting in the low equilibrium ($\bar{S}_0 = 0.20$) that experiences a grace-driven revival beginning in period 20.[1]

The transition exhibits the classic S-curve: slow at first (the low-S trap resists change), then accelerating (investment complementarity kicks in), then plateauing. GDP rises by approximately 30% over 60 periods. This is well short of the full steady-state gap of Table 4 (roughly 75%), and the difference is instructive rather than contradictory: over this horizon the average

[1]For visual clarity the transition exercise uses a smaller social multiplier ($\xi \approx 0.02$) and a smaller soul-production scale than the comparative-statics calibration of Table 3; this stretches the adjustment into a realistic multi-decade S-curve without altering its qualitative shape.

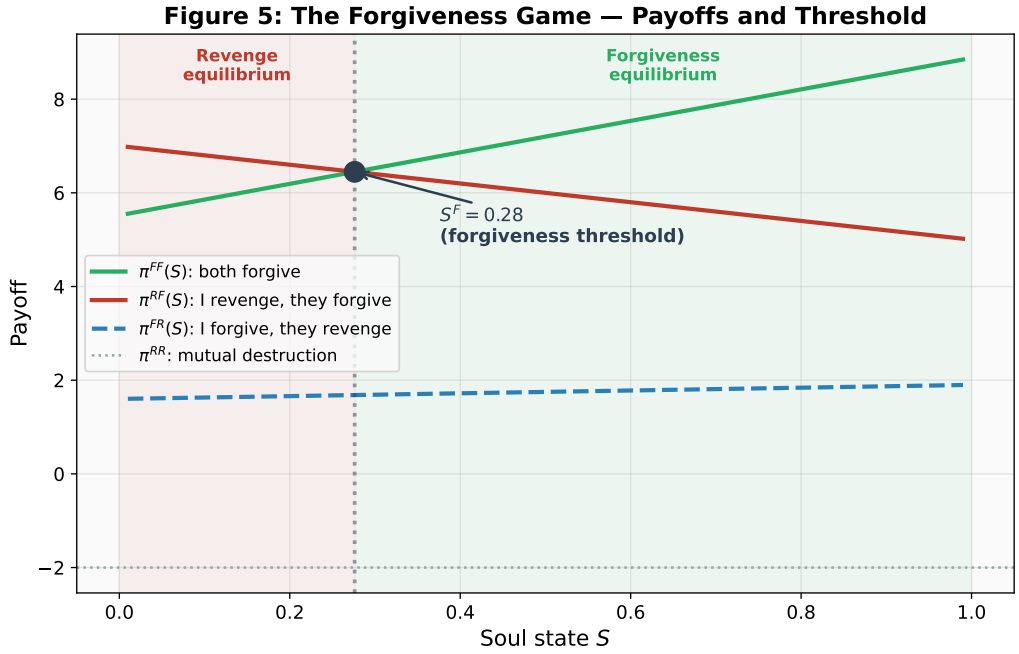


Figure 5: Payoffs in the forgiveness game. Above the threshold $S^F \approx 0.28$, forgiveness is strictly dominant and mutual forgiveness is the unique equilibrium. Below it, mutual forgiveness cannot be sustained: forgiving pays only if the other side forgives too, and every equilibrium leaves at least one group taking revenge (Proposition 5). The low-soul equilibrium ($\bar{S}_L \approx 0.20$) lies below this threshold.

soul state climbs only partway, from $\bar{S} = 0.20$ to about $\bar{S} = 0.53$, and human capital adjusts to greater patience with a lag, so the remaining gain accrues only as the transition completes. The pattern is consistent with post-revival economies: South Korea in the 1960s–90s (where Christianity grew from about 2% of the population in 1945 to roughly 30% by 1990 alongside the economic miracle), Rwanda in the 2000s–present, and the Wesleyan revival in 18th-century England that preceded the Industrial Revolution by decades.

7.7 Summary of simulation results

Table 4: Model outcomes at low and high soul states. The "high" column uses $\bar{S} \approx 0.70$, a reachable target just above the critical mass $\bar{S}^* \approx 0.65$; the model's high equilibrium proper sits higher still, at $\bar{S}_H \approx 0.85$.

Outcome	Low S (≈ 0.20)	High S (≈ 0.70)
Dishonesty rate θ	24%	9%
Giving/consumption ratio	∼5%	∼25%
Investment share of income	∼7%	∼25%
Measured TFP (fraction of true)	84%	94%
GDP (index, low $S = 100$)	100	∼175
Forgiveness game	Revenge in equilibrium	Forgiveness dominant

The magnitude should give pause. No tax incentive, no regulatory framework, no constitutional amendment can produce effects of this magnitude. The model says what Augustine said: the binding constraint on a society is not its institutions but the disposition of its citizens' hearts.

Figure 6: Transition Dynamics — A Grace-Driven Revival

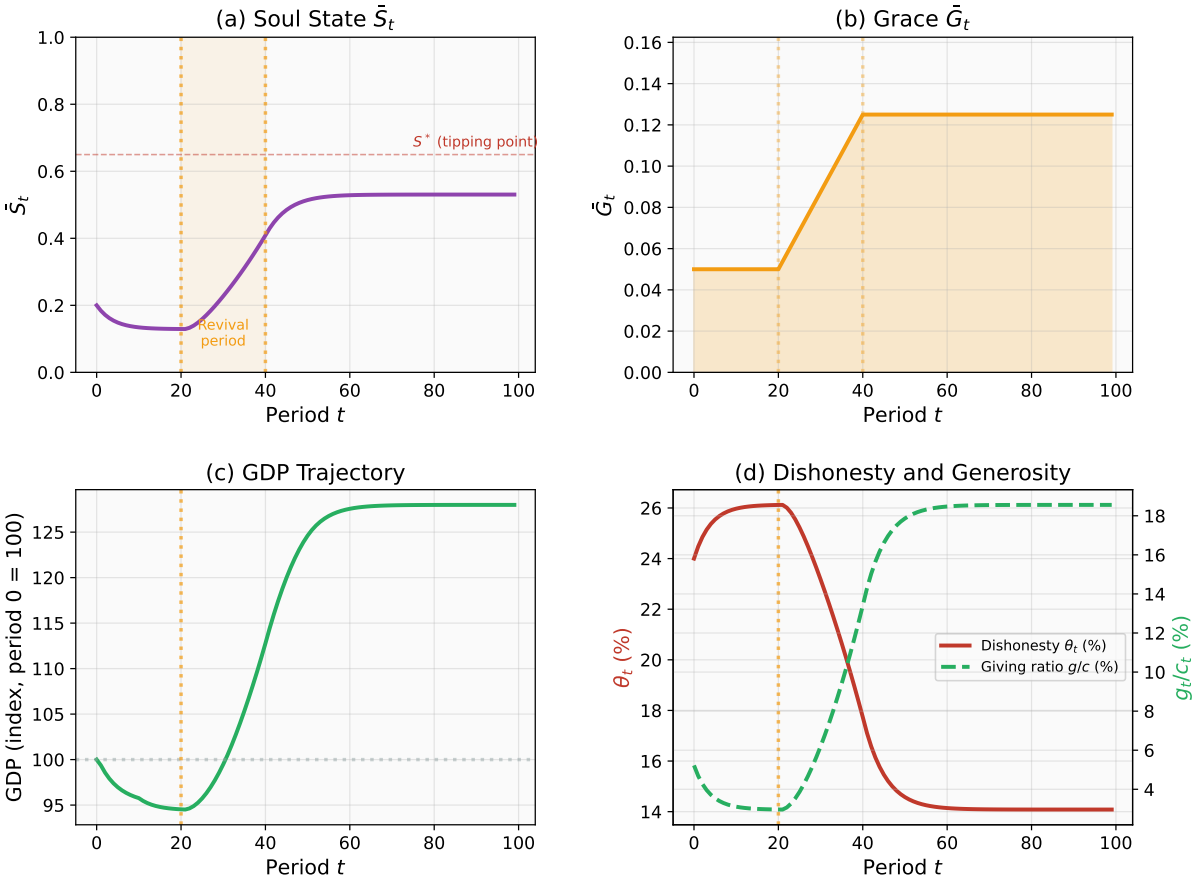


Figure 6: Transition dynamics following a grace-driven revival. The soul state follows an S-curve. GDP rises by approximately 30% over 60 periods. Dishonesty falls while the giving ratio rises. The transition takes a generation.

8 Nesting Result: Homo Economicus as a Special Case

The trichotomous model does not reject mainstream economics. It contains it.

Claim: In the no-grace limit ($\bar{G} = 0$ and $\varepsilon_t^G = 0$), with $S_0 = 0$, the soul stays at zero and the trichotomous agent reduces to the standard utility-maximizing agent of neoclassical economics.

Proof. Strictly, $S_t = 0$ for all t requires switching off grace: while $\bar{G} > 0$ the soul law of motion gives $S_{t+1} \geq \bar{G} - \rho\theta_t$ and the soul cannot remain identically zero. The pure standard model is therefore the *no-grace limit*, which is exactly the point: homo economicus is the agent God's grace has not (yet) reached. In that limit, with $S_0 = 0$ and $i^S = 0$, the soul stays at the floor $S = 0$ (bounded below at zero, since corrosion alone would push it negative). Setting $S_t = 0$:

- $\beta(0) = \underline{\beta}$: a constant discount factor, and $D_t = \underline{\beta}^t$.
- $\sigma(0) = \underline{\sigma}$: a constant curvature (standard CRRA utility).
- $\psi(0) = 0$: no weight on giving. The agent is purely self-interested.
- Honesty: with $S = 0$ the soul has no value ($\mu^S = 0$), so the corrosion cost vanishes and the honesty first-order condition collapses to $\nu_t = \lambda_t \Pi_t > 0$. The constraint $\theta_t \leq \bar{\theta}$ binds and the agent cheats maximally.
- Soul law of motion: $S_{t+1} = -\rho\theta_t$, floored at zero. With no grace and no investment, the soul never leaves zero.

The optimization problem reduces to:

$$\max \mathbb{E}_0 \sum_{t=0}^{T} \underline{\beta}^t \frac{c_t^{1-\underline{\sigma}}}{1-\underline{\sigma}} \quad \text{s.t.} \quad c_t + i_t^B + i_t^M + a_{t+1} \leq w_t M_t B_t + \bar{\theta}\Pi_t + (1+r_t)a_t \tag{15}$$

This is a textbook lifecycle model with human capital accumulation (Ben-Porath, 1967), health investment, and rent-seeking. The agent cheats to the maximum extent ($\theta = \bar{\theta}$) and has no reason not to, i.e., there is no internal moral constraint and no weight on the welfare of others. The soul investment term disappears because $\mu^S = 0$ when S has no value. □

This nesting result is important for two reasons. First, it means that any empirical test of the trichotomous model against the standard model is a test of $S > 0$ versus $S = 0$; it is a test of whether the soul matters, not whether the standard model is wrong. Second, it means that the trichotomous model is *at least as good* as the standard model in every application where the soul is irrelevant. We lose nothing by extending the framework and gain the ability to explain phenomena such as persistent corruption, the failure of institutional reform, the power of forgiveness, that the standard model cannot.

9 A Challenge to the Secular Economist

We address the skeptical reader directly. Five objections are likely.

9.1 "Norms can explain this without God"

You may prefer to call it norms rather than soul. Social norms, cultural capital, civic virtue are familiar terms and comfortable. Norms matter. Knack and Keefer (1997) show that trust predicts growth. Putnam (1993) demonstrates that civic engagement underpins institutional quality.

But one must answer: *where do norms come from?*

If you say evolution, we must take the sophisticated version of the answer. The crude version, "cooperation emerged because cooperative groups outcompeted selfish ones," is too loose to be useful. The serious accounts are kin selection (Hamilton), reciprocal altruism (Trivers), indirect reciprocity through reputation, and group selection. Grant that they work. Notice what they explain: *parochial* cooperation, that is, cooperation among kin, among repeated partners, among those who can observe and reward you. They explain why cooperation *persists*; they explain a great deal of moral *behavior*.

Two things they do not deliver. First, they do not bridge Hume's gap: an evolutionary account establishes that moral sentiments *exist* and confer fitness, but not that their dictates carry *normative authority*: "this trait was selected for" is an *is*, not an *ought*. We note the point and rest no weight on it: ours is a positive model, and nothing downstream depends on metaethics. The second gap is the one that carries the argument. Second, and more concretely, the prediction these accounts generate is that morality should be *parochial and reputation-bearing*. Yet the behavior our model and the Christian tradition place at the center is the opposite: costly, anonymous, one-shot generosity toward non-kin—and, harder still, toward enemies. A person who gives anonymously to a stranger he will never meet, who forgives an enemy who cannot reciprocate, is doing something the fitness calculus struggles to rationalize. The standard escape is the "big mistake" hypothesis: such generosity is a *misfiring* of instincts tuned for small kin-groups. But that concession is striking: it says impartial love of the stranger and the enemy is an evolutionary *error*. Our model, and the gospel it formalizes, treats it instead as the signature of a transformed soul. Ants cooperate beautifully. They do not forgive their enemies.

If you say culture, i.e., that norms are transmitted through families, schools, and institutions, you have pushed the question back one step. Where did the culture come from? Why do some cultures produce trust and others produce corruption? Acemoglu and Robinson (2012) argue that institutions are the answer, and North (1990) provides the foundational framework. But institutions are designed and operated by people, and the same institutional blueprint produces radically different outcomes in different societies. Culture is not an explanation. It is a description of the thing that needs explaining.

The most serious version of the culture answer deserves a direct response. Henrich (2020) argues that the medieval Western Church made the West psychologically peculiar: by dissolving the intensive kinship of clan and cousin-marriage, it produced individualism, analytic thinking, and the impersonal prosociality, or trust extended to strangers, on which markets and impersonal institutions run (Schulz et al., 2019). We take this as partial confirmation, not refutation. It is *Christianity*, specifically, that Henrich credits with reshaping the Western mind. But notice the boundary of the mechanism. Dissolving kinship can explain fair dealing with the stranger;

it does not explain love of the *enemy*, the costly forgiveness of one who cannot reciprocate and has no claim of blood or contract, which is the behavior we placed at the center a moment ago. And Henrich's account, like the rest of the culture literature, takes the Church's marriage program as a historical given: it offers no reason why impersonal prosociality should decay in some generations and revive in others. In our framework the kinship channel is one of the institutional shadows the soul casts, not a substitute for it.

The deepest version of the culture answer concedes religion's economic power while denying its object, and we owe it a direct reply. Norenzayan (2013) argues that belief in moralizing, punishing, all-seeing gods is itself a culturally evolved cooperation technology: groups whose members felt watched cooperated at scale, outcompeted their rivals, and spread, and the gods grew big because big gods were useful. On this account everything we attribute to grace can be attributed to *belief in* grace, and belief is endogenous to cultural evolution, which would reopen the bootstrap we closed in Section 3.3. The objection is observationally serious: a transformed heart and a sincerely believed-in watcher can look alike in a cross-section. They come apart in three places. First, the margin. Surveillance operates on the constraint set: a watcher, divine or human, raises the expected cost of cheating and so moves θ. It has no purchase on the other channels. Being watched does not make an agent patient, easily satisfied, or glad to give; yet the evidence connects belief to growth itself and to thrift, not only to honesty (Barro and McCleary, 2003; Guiso et al., 2003). A mechanism that is surveillance explains one channel in four; a mechanism that transforms S explains all of them together. Second, the duration. The experimental core of the big-gods evidence is priming: agents behave more prosocially when god-concepts are made salient, and the effect fades with the prime. That is what monitoring looks like. Transformation is what remains when the prime is gone, and the behavioral proxies we rely on (volunteering, giving, honesty when no one is watching) are measured unprimed. Third, manufacturability. If belief is a technology, deliberate manufacture should work: states that sponsored churches, mandated observance, and subsidized clergy should have produced durable belief and the cooperation that follows. The religion-market literature finds the opposite, since state regulation depresses belief and participation, and Europe's subsidized state churches preside over the continent's emptiest pews (Iannaccone, 1998; McCleary and Barro, 2006). Revival, where belief actually spreads, arrives unplanned and unpurchased, which is what an exogenous ε_t^G looks like in the data, and what a technology under human control does not.

If you say reason, i.e., that rational agents can derive moral norms from first principles, you face Hume's guillotine: you cannot derive an "ought" from an "is." Reason can tell you that cooperation is efficient. It cannot tell you that you *should* cooperate when defection is more profitable.

9.2 "The model is not falsifiable"

Grace is exogenous and stochastic. How can you test a model whose key variable is unobservable and uncontrollable?

We reply: the model generates *observable* predictions, but we must be precise about which observables, because a careless choice invites a fatal objection. The naive test would correlate *church attendance* with economic outcomes. We do not propose that test, and the model does

not license it. The predictions depend on the soul state S (genuine moral transformation), and Section 3.7 argued at length that nominal affiliation is not S. Indeed Barro and McCleary (2003) find that, conditional on belief, church *attendance* carries a *negative* sign: a society can sit in pews without being transformed. The correct proxies are *behavioral*: volunteering and charitable giving, honesty in incentivized experiments, contract compliance, marital stability, the believing-versus-belonging gap. With those proxies the model predicts, controlling for income, lower corruption (Proposition 1), higher savings (Proposition 2), more voluntary charity (Proposition 3), higher structural TFP (Proposition 4), and lower structural unemployment (Proposition 6).

We also concede something important: the static cross-section is a *weak* test. Consider the obvious challenge: secular Scandinavia and Switzerland are less corrupt, higher-saving, and higher-trust than the more religious United States. Does this not falsify the model? It would, if the model predicted that current attendance maps to current outcomes. It does not. But we must be careful about what it does predict, because the natural answer (that Scandinavia is running down an inherited stock of soul capital, which should therefore erode visibly across cohorts) is not the model's answer. At $\delta^S = 0.30$ (Table 3), an inherited stock alone would be gone within a handful of model periods. What sustains a secularizing Scandinavia in the model is not a stock running down but an *equilibrium holding*: the high steady state is locally stable, held in place by the social multiplier $\xi\bar{S}$, the grace floor $\bar{G}$, and the devotional investment that high-S agents still find worthwhile. Secularization, in the model, is a gradual downward shift of the replenishment curve in Figure 2; the high equilibrium drifts down a little and *persists*, until the curve falls far enough that the high crossing and the tipping point merge and the high basin ceases to exist. The predicted time path is therefore hysteresis: long stability, then rapid decline. A cliff, not a slope. And this matters because the data already speak. Generalized trust in Denmark has *risen*, not fallen, over four decades of deep secularization (Sønderskov and Dinesen, 2014). Under the stock-depreciation story that is a falsification. Under the equilibrium story it is what a society still inside the high basin looks like. Two cautions keep this honest. Attendance is a poor proxy for i^S in both directions (Section 3.7 refused full pews as evidence of high S; we must equally refuse empty pews as evidence of zero investment), and in a society where nearly everyone is still honest the social multiplier $\xi\bar{S}$ does exactly the sustaining work the model assigns it. The claim remains falsifiable, and in two directions. If effective devotional investment can fall without bound while the high equilibrium survives indefinitely, the replenishment mechanism is wrong. And when decline comes, the model requires it to be rapid, an S-curve downward, not a gentle cohort-by-cohort fade; a slow linear erosion stretched across generations would be evidence for the stock story and against the equilibrium structure at the heart of this paper.

Japan poses the sharper version of the same challenge, and it deserves to be named rather than evaded: a high-trust, low-corruption, high-growth society that was never Christianized, where the accumulated-Christianity defense we just gave for Scandinavia is unavailable. Confucian East Asia is the strongest standing counterexample to any claim that only Christianity produces the moral infrastructure of prosperity, and we do not make that claim: common grace is universal ($\bar{G} > 0$ for all societies, always), other traditions sustain real devotional investment i^S, and a strong social multiplier can hold a society at a high level of *coordinated* virtue. What the

model predicts is not that Japan should be dysfunctional but that its virtues should *decompose* differently: strong on the margins that monitoring, reputation, and reciprocity can produce, weak on the margins that require transformed preferences. That is what we observe. Japan's celebrated honesty operates inside a dense institutional web (the *kōban* police-box network and a lost-property law that entitles the finder to a statutory reward), and on the channels no watcher can reach, the picture inverts: Japan ranks near the bottom of international comparisons of charitable giving and volunteering, and last in the world on helping a stranger.[2] High coordination, low generosity: that is exactly the signature of a society running on the constraint-side technologies the standard model already understands, not on the preference transformation this paper is about. A generic trust index cannot see the difference. The four-channel decomposition can.

The model's real discriminating power is therefore *dynamic*, not cross-sectional. It predicts S-curve transitions in economic indicators following grace shocks; it predicts that institutional reform without spiritual transformation fails; it predicts the depreciation paths just described; and it predicts that the cross-country distribution of trust and corruption hides *two clusters* separated by an unstable threshold (Proposition 7). We have run the first test of that last prediction rather than promising it, and we report the results in full, including the part that goes against us (Figure 7). The naive version of the prediction fails. If every country shared the same parameters, the two equilibria would sit at the same two values everywhere and the pooled cross-section would show two visible humps. It does not: Hartigan's dip test finds no departure from unimodality for either generalized trust (dip $= 0.024$, $p = 0.93$, 117 countries, Integrated Values Surveys) or the Corruption Perceptions Index (dip $= 0.023$, $p = 0.77$, 182 countries); both densities are right-skewed with a single mode and a long high tail. The heterogeneous version survives. Countries do not share parameters (the grace floor, the social multiplier, and the honesty ceiling all differ), so the model predicts two *latent* clusters at country-specific locations that overlap when pooled, and that is what the data show: a two-component mixture is preferred over one for both distributions (BIC margins of 22.0 for trust and 5.4 for the corruption index), with the high cluster, trust near 40 percent and a corruption score near 66, carrying roughly a third of countries. A third measure, the World Bank's Worldwide Governance Indicators Control of Corruption estimate, repeats the structure on a longer window (1996–2024): no rejection of unimodality (dip $p = 0.87$, 207 countries), a two-component mixture preferred (BIC margin 4.0), components at -0.71 and $+0.90$ on its -2.5 to $+2.5$ scale, with 43 percent of countries in the high cluster. We add the necessary caution that the WGI and the CPI aggregate overlapping perception sources, so their agreement is consistency across aggregators, not independent replication. We grade this honestly: latent clusters are a weaker finding than visible bimodality, because strong skew alone can produce a preferred two-component fit, and the result earns the model consistency, not confirmation. One dynamic moment leans our way without settling anything: over a mean span of 26 years the trust panel shows no mean reversion, where classical measurement error alone would produce reversion, strongest at the extremes. A second exercise, in the distribution-dynamics spirit of Quah (1996), classifies countries into low, middle, and high bands using the mixture's own posterior (the middle band, where the posterior

[2] Charities Aid Foundation, *World Giving Index 2023*: Japan ranked 139th of 142 countries overall, and 142nd on the "helped a stranger" measure.

lies between 0.25 and 0.75, is the data's estimate of the watershed zone) and counts band-to-band transitions. The raw pattern looks like a ridge: over a mean span of 26 years only 24 percent of middle-band trust countries stay put, against 83 percent in the low band and 60 percent in the high, middle exits split almost evenly in the two directions, and in the corruption panel not one country jumps directly between the poles in a decade. But we benchmark before we celebrate. Under a permutation null that hands each country a change drawn from the pooled change distribution irrespective of its starting level, much of that ranking is reproduced by band widths alone: the middle band is narrow, so even level-blind dynamics empty it. What survives the null is the economically central part. The low band holds its countries more tightly than mechanics can explain (83 percent observed against 72 percent under the null, $p = 0.03$): the low equilibrium behaves like a trap in precisely the way Proposition 7 requires. And the symmetric two-way leakage from the middle is inconsistent with reversion toward the distribution's single mode, which sits inside the low band and would pull middle exits downward. The corruption window is too short to distinguish anything, and the high trust band churns more than the null. On inspection the churn is concentrated at the basin's edge: persistence rises with distance from the watershed in all three datasets, and in the World Bank panel, whose window extends to twenty-four years, every one of the thirty-one countries in the deep half of the high band stayed, while all thirteen exits came from the shallow half. We discipline this the way we disciplined the stay rates, and the discipline humbles us: a level-blind change process also concentrates exits at the edge, because deep countries need large moves to exit at all, and the observed concentration does not significantly beat that benchmark ($p \approx 0.11$ at its strongest). The deep basins are quiet, as the model requires, but quiet is also what small noise far from a boundary produces. Two countries deserve naming, because they are among the rare full crossers and they are the two this paper argues from (Section 6.2). Rwanda began the window at -0.80 on the World Bank scale, the low cluster's center, and stands today at $+0.63$, across the watershed. South Africa began at $+0.87$, inside the high band, and has slid to -0.19, into the low basin. The paper's two cases are, in the World Bank's own panel, mirror-image crossings of the same threshold in opposite directions. With fifteen to seventeen middle-band countries and a handful of named crossings, the grade remains the one we gave the mixture: consistency, not confirmation. The empirical literature remains supportive on the levels: Barro and McCleary (2003) find religious belief predicts growth; Guiso et al. (2006) show denomination affects economic attitudes; Putnam (1993) documents two sharply separated clusters within a single country; and the contrast between Rwanda and South Africa, two cases rather than a controlled comparison and no substitute for systematic evidence, is at least suggestive of the model's key mechanism. A full empirical test, structural estimation against these dynamic moments, is the necessary next step, and we flag it as such rather than claiming it is already done.[3]

[3]Data and replication code: Our World in Data compilations of the Integrated Values Surveys trust share (latest wave per country), Transparency International's Corruption Perceptions Index (latest year, post-2012 methodology), and the World Bank's Worldwide Governance Indicators Control of Corruption estimate (1996–2024, real countries only); `dip_test_bimodality.py`, `bimodality_dynamics.py`, `transition_matrix.py`, and `wgi_corruption.py` in the companion materials.

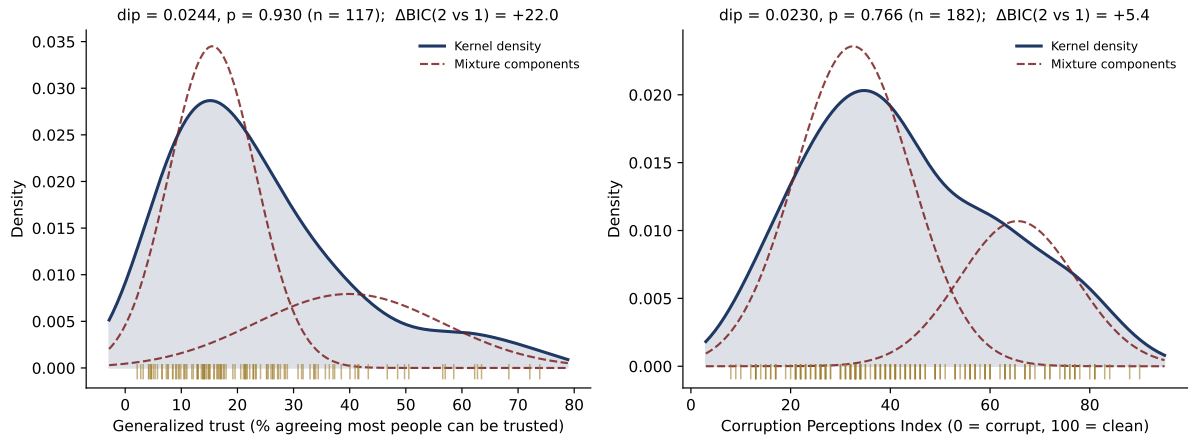


Figure 7: No visible bimodality, but two latent clusters. Kernel densities (solid) of the latest cross-country distributions of generalized trust (left) and the Corruption Perceptions Index (right), with fitted two-component Gaussian mixtures (dashed). Hartigan's dip test cannot reject unimodality in either panel; a two-component mixture is nonetheless preferred by BIC in both, with the high cluster carrying roughly a third of countries. Rug marks: individual countries.

9.3 "You have a reverse causality problem"

Does higher S cause prosperity, or does prosperity cause higher S? This is a legitimate concern. Three responses.

First, the *timing* evidence favors our direction. Woodberry (2012) demonstrates that conversionary Protestant missions predict democracy, literacy, and economic development decades later, and the missions preceded the development, not the reverse. South Korea's Christian population grew from roughly 2 percent in 1945 to about a fifth of the population by 1985 and nearly 30 percent by 2005. We state the timing carefully, because the bulk of that growth ran *alongside* the boom rather than before it: when the first five-year plan launched in 1962, Christians were perhaps one in twenty Koreans. What the record will not support is the reverse story, that prosperity manufactured the faith: the missionary schools and the literacy they spread date from the late nineteenth century, the Pyongyang revival from 1907, and the church grew in what was then one of the poorest countries on earth. Korea is Woodberry's mechanism in miniature, missions first, human capital next, growth last. The Wesleyan revival in 18th-century England preceded the Industrial Revolution by decades. Rwanda's Christian revival preceded, not followed, its 7–8% GDP growth.

Second, the model generates *distinguishing predictions* that the reverse-causality story cannot match. If prosperity causes religiosity, we should see rich countries becoming more religious over time. The opposite is happening in Western Europe. If prosperity causes forgiveness, we should see rich societies forgiving historical wrongs more easily. We do not. Wealthy, secular Western Europe has not forgiven more readily than comparatively poorer, Christian Rwanda. The model's specific predictions about forgiveness equilibria, voluntary giving replacing state redistribution, and S-curve transitions following grace shocks have no natural reverse-causality explanation.

Third, the nesting result provides a *falsification criterion*. If $S = 0$ and the standard model is sufficient, then institutional quality alone should predict economic outcomes after controlling for

capital, labor, and technology. The persistent failure of well-designed institutions in low-trust environments such as South Africa's Constitution, the Chapter 9 institutions, an independent central bank is evidence that the standard model is incomplete. Something is missing. We identify that something as the soul.

9.4 "You are smuggling theology into economics"

In a sense, yes, and the more useful observation is that setting $S = 0$ is not a neutral default either. It is itself a substantive assumption about human nature, no more observed than its negation. Every model makes metaphysical commitments. The standard model assumes that the agent is fully described by preferences over consumption and leisure, that these preferences are fixed, and that moral considerations enter only through external constraints. These are not value-neutral assumptions; they are modeling choices, and the soul is excluded by convention rather than by evidence. We do not charge the standard modeler with bad faith. We ask only that the choice be recognized as a choice, so that $S = 0$ and $S > 0$ can be compared on the evidence rather than one of them being treated as the costless default.

We propose a more general framework. If you believe $S = 0$, our model reduces to the standard framework and one loses nothing. If $S > 0$, the model explains phenomena that yours cannot. The question is not whether we are smuggling theology into economics. The question is whether the data is better explained with or without the soul.

9.5 "This is a god-of-the-gaps argument"

The objection runs as follows: you cannot explain why some societies flourish and others decay, so you insert God into the residual and call it grace. This is the "god of the gaps"—an argument from ignorance dressed in equations.

We take this objection seriously. The god-of-the-gaps fallacy is real: attributing the currently unexplained to divine action, only to retreat when science advances, is bad theology and worse science. But that is not what we do. Our argument is not "we cannot explain the residual, therefore God." Our argument is: here is a more general model that nests the standard one, generates additional testable predictions, and fits the data better on specific margins. The nesting result (Section 8) is the key: if $S = 0$ is sufficient, reject our model. We have given you the falsification criterion. A god-of-the-gaps argument provides no such criterion.

Consider the parallel. When Solow (1956) attributed the majority of growth to an unexplained residual and called it "technology," no one accused him of a "technology of the gaps" argument. When Real Business Cycle theorists modeled productivity as an exogenous stochastic process, no one demanded that they first explain where technology shocks come from before being allowed to study their consequences. We model grace as an exogenous stochastic process and study its consequences. So far the methodology is parallel, but it is not yet *identical*, and the difference is the whole point. What earned the RBC program its license was not the act of positing an exogenous shock; it was the discipline imposed afterward. Kydland and Prescott (1982) did not claim to *explain* business cycles. What they did was ask a precise quantitative question: how much of the variance of output around trend can persistent technology shocks reproduce? They hung their hat on *matching the data*. To claim the same license, we must do

the same: identify, concretely, what current models fail to match in the data, not in theology, and show that ours matches it. We have named the candidate facts: the bimodality of the cross-country trust and corruption distribution, the failure of institutional reform to move outcomes, and the S-curve timing of post-revival transitions (Section 9). The honest status of this paper is that it supplies the model and the discriminating predictions; the structural estimation that would complete the parallel remains to be done. We claim methodological *legitimacy* for an exogenous spiritual residual, not a finished empirical victory.

But we go further than methodological parity. We make a specific claim about the *source* and *mechanism* of grace that no secular residual can match. The claim is this: grace enters the world through the cross of Christ.

The cross is the model's deepest exogenous intervention. In Christ, God enters the system from outside, not as a metaphor, not as a cultural force, but as a historical person who lived, died under Roman execution, and rose from the dead. The crucifixion is the event in which the just penalty for human rejection of God (the very rejection that produces the low-S equilibrium) is borne by God himself. The resurrection is the demonstration that the penalty has been fully paid. And the gift that flows from this event (grace, unearned and unearnable) is what makes spiritual transformation possible. Christ died so that all who believe in Him would have eternal life. And even the faith to believe is not the agent's own doing: "For by grace you have been saved through faith. And this is not your own doing; it is the gift of God, not a result of works, so that no one may boast" (Ephesians 2:8–9). Grace is sovereign from first to last. The decision to bestow it was made before the foundation of the world (Ephesians 1:4). Grace is exogenous not merely contemporaneously but *eternally*.

The cross answers both objections simultaneously:

1. **It is not a gap-filler.** The cross is a specific historical claim, open to historical evidence in a way no gap-filler is, and Paul himself stated the falsification condition: if Christ did not rise, the entire mechanism collapses (1 Corinthians 15:14). A god-of-the-gaps argument retreats from evidence. The Christian claim *invites* it. We keep the two layers distinct: the cross is the claimed *source* of grace, a matter of historical evidence; the model's testable consequences are economic, and they are the ones catalogued in Section 9.2.

2. **Grace is universal *and* particular.** "For God so loved the *world*" (John 3:16); this love is expressed in common grace, which sustains all of creation. In the model, $\bar{G} > 0$ for all agents, always. No society is abandoned. The rain falls on the just and the unjust. But *effectual* grace, the transformative shock ε_t^G that regenerates the heart, is particular. Christ died so that all who believe in Him would have eternal life, and even the faith to believe is itself a gift of grace (Ephesians 2:8–9), granted to those whom God chose before the foundation of the world (Ephesians 1:4). The agent cannot generate ε_t^G from within. This is not a restriction on God's generosity. It is the deepest expression of His sovereignty: that salvation, from first to last, is His work and not ours.

This leads to a final point about the low-S trap. Common grace ($\bar{G}$) keeps even the lowest-S society from total collapse; it sustains a residual conscience, a minimal moral order, a dim awareness that corruption is wrong even where it is practiced. But common grace alone cannot

lift a society across the threshold S^*. That requires effectual grace: the stochastic shocks ε_t^G that transform individual hearts, producing agents whose patience, honesty, generosity, and satiation generate the social multiplier $\xi \bar{S}_t$ that pulls others upward. The low-S equilibrium persists not because God withholds common grace (He does not) but because effectual grace has not yet reached the critical mass. "Light has come into the world, and people loved the darkness rather than the light" (John 3:19). And the transition to the high equilibrium requires what the model predicts: a critical mass of agents in whom effectual grace has worked, whose transformed behavior generates the social multiplier that pulls the society across the threshold. This is, in economic language, what the New Testament calls the Church.

10 Conclusion

We have presented a formal economic model in which the agent is composed of body, mind, and soul. The soul transforms the utility function through four channels: patience, satiation, generosity, and honesty. Grace, an exogenous gift from God, provides the mechanism for spiritual transformation that human effort alone cannot produce. The model nests standard homo economicus as the special case $S = 0$ and generates seven testable propositions connecting the soul state to observable economic outcomes.

The model's deepest claim is also its most controversial: that the binding constraint on human flourishing is not capital, not technology, not institutions, but the disposition of the human soul. Augustine said it in the fifth century. We have said it in the language of dynamic optimization.

This is not a paper that will satisfy everyone. The secular economist will object to the exogeneity of grace. The theologian may object to the formalization of the sacred. The policy maker may wonder what to *do* with a model whose key variable cannot be legislated.

To the economist, we say: you already model exogenous variables you cannot fully explain. Grace deserves the same treatment.

To the theologian, we say: formalization is not reduction. The model does not claim to capture the fullness of God's grace. It claims only to capture its *economic shadow*, or the way it transforms behavior, incentives, and outcomes in the measurable world.

To the policy maker, we say: you cannot legislate grace, but you can stop obstructing it. Protect religious freedom. Support churches and faith communities. Design institutions that reward virtue rather than gaming. And understand that the deepest reforms, the ones that actually transform a society, come not from Pretoria or Washington or Brussels, but from the human heart, touched by a power that no economist can endogenize.

Augustine ended *De Civitate Dei* with a vision of the eternal Sabbath: the rest that comes when the pilgrim city finally reaches its home. We end with the more modest hope that the tools of economics, honestly applied, can point in the same direction.

Soli Deo Gloria.

References

Augustine of Hippo (426). *De Civitate Dei* [The City of God]. Translated by Henry Bettenson, Penguin Classics, 2003.

Barro, R. J. and McCleary, R. M. (2003). Religion and Economic Growth across Countries. *American Sociological Review*, 68(5):760–781.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*, 75(4):352–365.

Guiso, L., Sapienza, P., and Zingales, L. (2006). Does Culture Affect Economic Outcomes? *Journal of Economic Perspectives*, 20(2):23–48.

Iannaccone, L. R. (1998). Introduction to the Economics of Religion. *Journal of Economic Literature*, 36(3):1465–1495.

Kierkegaard, S. (1849). *The Sickness Unto Death*. Translated by Howard V. Hong and Edna H. Hong, Princeton University Press, 1980.

Knack, S. and Keefer, P. (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics*, 112(4):1251–1288.

Putnam, R. D. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press.

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics*, 70(1):65–94.

Stark, R. (2005). *The Victory of Reason: How Christianity Led to Freedom, Capitalism, and Western Success*. Random House.

Weber, M. (1905). *The Protestant Ethic and the Spirit of Capitalism*. Translated by Talcott Parsons, Scribner, 1930.

Woodberry, R. D. (2012). The Missionary Roots of Liberal Democracy. *American Political Science Review*, 106(2):244–274.

Acemoglu, D. and Robinson, J. A. (2012). *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. Crown Business, New York.

North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press.

McCleary, R. M. and Barro, R. J. (2006). Religion and Economy. *Journal of Economic Perspectives*, 20(2):49–72.

Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. Free Press, New York.

Tabellini, G. (2010). Culture and Institutions: Economic Development in the Regions of Europe. *Journal of the European Economic Association*, 8(4):677–716.

Alesina, A. and Giuliano, P. (2015). Culture and Institutions. *Journal of Economic Literature*, 53(4):898–944.

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *Economic Journal*, 100(401):464–477.

Becker, G. S. and Mulligan, C. B. (1997). The Endogenous Determination of Time Preference. *Quarterly Journal of Economics*, 112(3):729–758.

Bénabou, R. and Tirole, J. (2011). Identity, Morals, and Taboos: Beliefs as Assets. *Quarterly Journal of Economics*, 126(2):805–855.

Mazar, N., Amir, O., and Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, 45(6):633–644.

Becker, S. O. and Woessmann, L. (2009). Was Weber Wrong? A Human Capital Theory of Protestant Economic History. *Quarterly Journal of Economics*, 124(2):531–596.

Feddersen, T. and Sandroni, A. (2006). A Theory of Participation in Elections. *American Economic Review*, 96(4):1271–1282.

Guiso, L., Sapienza, P., and Zingales, L. (2003). People's Opium? Religion and Economic Attitudes. *Journal of Monetary Economics*, 50(1):225–282.

Henrich, J. (2020). *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous*. Farrar, Straus and Giroux, New York.

Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., and Henrich, J. (2019). The Church, Intensive Kinship, and Global Psychological Variation. *Science*, 366(6466):eaau5141.

Norenzayan, A. (2013). *Big Gods: How Religion Transformed Cooperation and Conflict*. Princeton University Press.

Kydland, F. E. and Prescott, E. C. (1982). Time to Build and Aggregate Fluctuations. *Econometrica*, 50(6):1345–1370.

Longman, T. (2010). *Christianity and Genocide in Rwanda*. Cambridge University Press.

Tutu, D. (1999). *No Future Without Forgiveness*. Doubleday, New York.

Bellah, R. N. (1957). *Tokugawa Religion: The Values of Pre-Industrial Japan*. Free Press, Glencoe, IL.

Botticini, M. and Eckstein, Z. (2012). *The Chosen Few: How Education Shaped Jewish History, 70–1492*. Princeton University Press, Princeton, NJ.

Blum, U. and Dudley, L. (2001). Religion and Economic Growth: Was Weber Right? *Journal of Evolutionary Economics*, 11(2):207–230.

Eisenstadt, S. N., editor (1968). *The Protestant Ethic and Modernization: A Comparative View*. Basic Books, New York.

Samuelsson, K. (1993). *Religion and Economic Action: The Protestant Ethic, the Rise of Capitalism, and the Abuses of Scholarship*. Translated by E. Geoffrey French. University of Toronto Press, Toronto. First published 1957.

Sønderskov, K. M. and Dinesen, P. T. (2014). Danish Exceptionalism: Explaining the Unique Increase in Social Trust over the Past 30 Years. *European Sociological Review*, 30(6):782–795.

Butner Jr., D. G. (2024). *Work Out Your Salvation: A Theology of Markets and Moral Formation*. Fortress Press, Minneapolis.

Quah, D. T. (1996). Twin Peaks: Growth and Convergence in Models of Distribution Dynamics. *Economic Journal*, 106(437):1045–1055.

A Derivations

A.1 The Lagrangian

The agent maximizes:

$$V_0 = \mathbb{E}_0 \sum_{t=0}^{T} D_t \, u(c_t, g_t;\, S_t), \qquad D_t \equiv \prod_{s=1}^{t} \beta(S_s), \quad D_0 \equiv 1$$

subject to:

$$\begin{aligned}
c_t + g_t + i_t^B + i_t^M + i_t^S + a_{t+1} &= w_t M_t B_t + \theta_t \Pi_t + (1 + r_t) a_t && \text{(Budget)} \\
0 \le \theta_t &\le \bar{\theta}(1 - S_t) && \text{(Honesty)} \\
B_{t+1} &= (1 - \delta_t^B) B_t + \phi^B(i_t^B) && \text{(Body)} \\
M_{t+1} &= (1 - \delta^M) M_t + \phi^M(i_t^M, M_t) && \text{(Mind)} \\
S_{t+1} &= (1 - \delta^S) S_t + \phi^S(i_t^S) + \xi \bar{S}_t + \mathcal{G}_t - \rho\, \theta_t && \text{(Soul)}
\end{aligned}$$

with a_0 given, $a_{T+1} \ge 0$, and the soul transition understood as truncated to $[0, 1]$ (equation (7)); all derivative-based steps below are for interior S.

The Lagrangian is:

$$\begin{aligned}
\mathcal{L} = \mathbb{E}_0 \sum_{t=0}^{T} \Bigg\{ & D_t \left[\frac{c_t^{1-\sigma(S_t)}}{1 - \sigma(S_t)} + \psi(S_t) \frac{g_t^{1-\sigma(S_t)}}{1 - \sigma(S_t)} \right] \\
& - \lambda_t \left[c_t + g_t + i_t^B + i_t^M + i_t^S + a_{t+1} - w_t M_t B_t - \theta_t \Pi_t - (1 + r_t) a_t \right] \\
& - \nu_t \left[\theta_t - \bar{\theta}(1 - S_t) \right] + \kappa_t\, \theta_t \\
& - \mu_t^B \left[B_{t+1} - (1 - \delta_t^B) B_t - \phi^B(i_t^B) \right] \\
& - \mu_t^M \left[M_{t+1} - (1 - \delta^M) M_t - \phi^M(i_t^M, M_t) \right] \\
& - \mu_t^S \left[S_{t+1} - (1 - \delta^S) S_t - \phi^S(i_t^S) - \xi \bar{S}_t - \mathcal{G}_t + \rho\, \theta_t \right] \Bigg\}
\end{aligned} \tag{16}$$

where $D_t \equiv \prod_{s=1}^{t} \beta(S_s)$ is the cumulative discount factor of equation (1), $\nu_t \ge 0$ is the multiplier on the moral ceiling, and $\kappa_t \ge 0$ the multiplier on the lower bound $\theta_t \ge 0$, with complementary slackness $\nu_t[\bar{\theta}(1 - S_t) - \theta_t] = 0$ and $\kappa_t \theta_t = 0$.

Multiplier convention. The multipliers above are in *present value* (units of time-0 utility). Define the current-value counterparts by dividing by the cumulative discount factor,

$$\tilde{\lambda}_t \equiv \frac{\lambda_t}{D_t}, \quad \tilde{\nu}_t \equiv \frac{\nu_t}{D_t}, \quad \tilde{\kappa}_t \equiv \frac{\kappa_t}{D_t}, \quad \tilde{\mu}_t^j \equiv \frac{\mu_t^j}{D_t} \;\; (j = B, M, S).$$

Every first-order and co-state condition below is stated in current value, and only in current value; mixing the two conventions is how sign and timing errors creep in, so we refuse the mixture. Because $D_{t+1}/D_t = \beta(S_{t+1})$, the conversion is mechanical: a present-value condition linking t to $t + 1$ acquires exactly one factor $\beta(S_{t+1})$ when written in current value. Conditions

that live entirely within period t (the static ones) are form-invariant. The main text quotes the current-value conditions with the tildes suppressed for readability.

A.2 First-order conditions

Consumption:

$$\frac{\partial \mathcal{L}}{\partial c_t} = D_t \cdot c_t^{-\sigma(S_t)} - \lambda_t = 0 \quad \Longrightarrow \quad c_t^{-\sigma(S_t)} = \tilde{\lambda}_t$$

Giving:

$$\psi(S_t) \cdot g_t^{-\sigma(S_t)} = \tilde{\lambda}_t$$

Setting equal with the consumption FOC and rearranging:

$$\left(\frac{g_t}{c_t}\right)^{\sigma} = \psi(S_t) \quad \Longrightarrow \quad \frac{g_t}{c_t} = \psi(S_t)^{1/\sigma(S_t)}$$

The financial asset: a_{t+1} appears in the budget constraint at t (with multiplier λ_t) and at $t+1$ (through the return $(1+r_{t+1})a_{t+1}$, with multiplier λ_{t+1}):

$$\frac{\partial \mathcal{L}}{\partial a_{t+1}} = -\lambda_t + \mathbb{E}_t\left[\lambda_{t+1}(1+r_{t+1})\right] = 0 \quad \Longrightarrow \quad \tilde{\lambda}_t = \mathbb{E}_t\left[\beta(S_{t+1})(1+r_{t+1})\,\tilde{\lambda}_{t+1}\right]$$

using $\lambda_{t+1} = D_{t+1}\tilde{\lambda}_{t+1} = D_t\,\beta(S_{t+1})\tilde{\lambda}_{t+1}$. This is the intertemporal condition the Euler equation is built from; without the asset there is no such margin and no Euler equation.

Investment in body, mind, and soul:

$$\tilde{\lambda}_t = \tilde{\mu}_t^B\,\phi^{B\prime}(i_t^B) \tag{17}$$

$$\tilde{\lambda}_t = \tilde{\mu}_t^M \frac{\partial \phi^M}{\partial i_t^M} \tag{18}$$

$$\tilde{\lambda}_t = \tilde{\mu}_t^S\,\phi^{S\prime}(i_t^S) \tag{19}$$

At the optimum, the marginal cost of investment ($\tilde{\lambda}_t$) equals the marginal product times the shadow value for each type of capital. These conditions are static, so they hold identically in either multiplier convention.

Honesty: differentiating $\mathcal{L}$ with respect to θ_t, the budget term contributes $+\lambda_t\Pi_t$, the ceiling contributes $-\nu_t$, the floor contributes $+\kappa_t$, and the soul constraint, through the corrosion term $+\rho\theta_t$ inside its bracket, contributes $-\mu_t^S\rho$. In current value:

$$\tilde{\lambda}_t\Pi_t - \tilde{\mu}_t^S\,\rho - \tilde{\nu}_t + \tilde{\kappa}_t = 0, \qquad \tilde{\kappa}_t\,\theta_t = 0, \qquad \tilde{\nu}_t\left[\bar{\theta}(1-S_t) - \theta_t\right] = 0.$$

The Lagrangian is linear in θ_t, so the solution is generically at a corner and the multipliers sort the cases. If $\tilde{\mu}_t^S\rho > \tilde{\lambda}_t\Pi_t$ (the spiritual cost of corrosion exceeds the rents on offer), then $\theta_t = 0$, $\tilde{\nu}_t = 0$, and the floor multiplier $\tilde{\kappa}_t = \tilde{\mu}_t^S\rho - \tilde{\lambda}_t\Pi_t > 0$ closes the first-order condition at the corner; the agent chooses honesty freely. Note that without $\tilde{\kappa}_t$ the condition could not hold there at all: the equality $\tilde{\lambda}\Pi = \tilde{\nu} + \tilde{\mu}^S\rho$ with $\tilde{\nu} = 0$ would demand $\tilde{\lambda}\Pi = \tilde{\mu}^S\rho$ exactly, which is a knife-edge, not a region. If instead $\tilde{\mu}_t^S\rho < \tilde{\lambda}_t\Pi_t$, the ceiling binds: $\theta_t = \bar{\theta}(1-S_t)$, $\tilde{\kappa}_t = 0$, and $\tilde{\nu}_t = \tilde{\lambda}_t\Pi_t - \tilde{\mu}_t^S\rho > 0$. Without the corrosion term ($\rho = 0$) one recovers $\tilde{\nu}_t = \tilde{\lambda}_t\Pi_t > 0$ identically,

so the ceiling would always bind, the degenerate case the corrosion mechanism is introduced to avoid.

A.3 Co-state equations

Shadow value of body:

$$\tilde{\mu}_t^B = \beta(S_{t+1}) \left[\tilde{\lambda}_{t+1} w_{t+1} M_{t+1} + \tilde{\mu}_{t+1}^B (1 - \delta_{t+1}^B) \right] \tag{20}$$

Shadow value of mind:

$$\tilde{\mu}_t^M = \beta(S_{t+1}) \left[\tilde{\lambda}_{t+1} w_{t+1} B_{t+1} + \tilde{\mu}_{t+1}^M \left((1 - \delta^M) + \frac{\partial \phi^M}{\partial M_{t+1}} \right) \right] \tag{21}$$

The single factor $\beta(S_{t+1})$ in front of each bracket is the current-value convention at work: in present value the same equations read $\mu_t^B = \lambda_{t+1} w_{t+1} M_{t+1} + \mu_{t+1}^B (1 - \delta_{t+1}^B)$ with no explicit discount factor, because the present-value multipliers already carry D_{t+1} inside them. Note also that the shadow values of body and mind depend on legitimate earnings $w_t M_t B_t$ (not reduced by θ). The honesty choice affects income through the $\theta_t \Pi_t$ rent term, not through the productivity of body and mind. The term $\partial \phi^M / \partial M$ is the self-productivity of human capital (Ben-Porath effect): existing knowledge helps acquire new knowledge.

Shadow value of soul. S_{t+1} enters the problem in five places: the soul law at t (giving $-\tilde{\mu}_t^S$ after conversion); the discount factor $\beta(S_{t+1})$, which multiplies the *entire* continuation value from $t+1$ onward, since every D_s with $s \geq t+1$ contains it; the period-$(t{+}1)$ utility through $\sigma(S_{t+1})$ and $\psi(S_{t+1})$; the honesty ceiling at $t+1$; and the soul law at $t+1$ through $(1 - \delta^S) S_{t+1}$. Collecting the terms in current value:

$$\tilde{\mu}_t^S = \beta'(S_{t+1})\, V_{t+1} + \beta(S_{t+1}) \left[\frac{\partial u_{t+1}}{\partial \sigma} \frac{\partial \sigma}{\partial S} + \frac{\partial u_{t+1}}{\partial \psi} \frac{\partial \psi}{\partial S} - \tilde{\nu}_{t+1} \bar{\theta} + \tilde{\mu}_{t+1}^S (1 - \delta^S) \right] \tag{22}$$

where $V_{t+1} \equiv \mathbb{E}_{t+1} \sum_{s=t+1}^{T} (D_s / D_{t+1})\, u_s$ is the current-value continuation utility. Two details of this equation repay attention, because earlier drafts got both wrong. First, the *timing* of the patience term: $\beta(S_{t+1})$ scales everything from $t+1$ onward, so its derivative multiplies V_{t+1}, the whole remaining life, not merely the value from $t+2$. Second, the *sign* of the honesty term. The ceiling enters the Lagrangian as $-\nu_{t+1}[\theta_{t+1} - \bar{\theta}(1 - S_{t+1})]$; differentiating with respect to S_{t+1} gives $-\nu_{t+1}\bar{\theta}$, and no rearrangement changes that. Economically: when the ceiling binds, a higher soul state tightens it and removes a privately profitable option from the feasible set, which *lowers* the agent's private value. That is the envelope theorem, and it is also the theology: conscience costs something precisely while temptation is still alive. The term vanishes when the agent chooses honesty freely ($\tilde{\nu}_{t+1} = 0$) and is deepest at low S, so the private return to the soul is weakest exactly where society is trapped; the positive social value of that same honesty is Proposition 4, and the wedge between the two is the point. If one insists on a positive intrinsic value of integrity to the agent, it must be placed in preferences or the continuation value explicitly; it cannot be conjured by flipping the sign of a Kuhn–Tucker term.

A.4 Steady-state conditions

Steady states belong to the stationary version of the problem (Remark 4.6): constant δ^B and an infinite horizon or, equivalently, a stationary population of overlapping generations. Within that version, in steady state ($B_{t+1} = B_t$, $M_{t+1} = M_t$, $S_{t+1} = S_t$, $a_{t+1} = a_t$, $c_{t+1} = c_t$):

$$\delta^B B^* = \phi^B(i^{B*}) \tag{23}$$

$$\delta^M M^* = \phi^M(i^{M*}, M^*) \tag{24}$$

$$\delta^S S^* = \phi^S(i^{S*}) + \bar{G} - \rho\,\theta^* \tag{25}$$

(the soul condition is written for the individual, holding $\bar{S}$ fixed; the equilibrium version replaces δ^S with $\delta^S - \xi$ once $S^* = \bar{S}^*$). The asset margin adds the familiar condition $\beta(S^*)(1 + r) = 1$ from the Euler equation with constant consumption: a higher steady-state soul, being more patient, is consistent with a lower interest rate.

The grace floor: if $i^{S*} = 0$ and the agent also does not cheat ($\theta^* = 0$), then $S^*_{\min} = \bar{G}/\delta^S$ for the individual agent. A fully negligent agent who also cheats to the moral limit, $\theta^* = \bar{\theta}(1 - S^*)$, has the lower floor $S^*_{\min} = (\bar{G} - \rho\bar{\theta})/(\delta^S - \rho\bar{\theta})$. In equilibrium, when the social multiplier is included ($S^* = \bar{S}^*$), δ^S is replaced by $\delta^S - \xi$ throughout, raising the floor: the community sustains a baseline soul state above what any isolated individual could maintain. Even a completely negligent agent maintains a non-zero soul state, because God's grace persists, and under the calibration it more than offsets the corrosion of habitual dishonesty.

A.5 The consumption Euler equation

Substituting the consumption first-order conditions at t and $t + 1$ into the asset condition $\tilde{\lambda}_t = \beta(S_{t+1})(1 + r_{t+1})\tilde{\lambda}_{t+1}$:

$$c_t^{-\sigma(S_t)} = \beta(S_{t+1})(1 + r_{t+1})\, c_{t+1}^{-\sigma(S_{t+1})} \tag{26}$$

Because lifetime utility is discounted by the *product* $D_t = \prod_{s \le t} \beta(S_s)$ rather than by β^t, the ratio $D_{t+1}/D_t = \beta(S_{t+1})$ is exact, and this Euler equation holds exactly even when the soul state varies over the life cycle: no constant-S approximation is required at this step. Two departures from the standard Euler equation remain: (1) the discount factor $\beta(S_{t+1})$ depends on the soul, a more spiritual agent is more patient; (2) the curvature $\sigma(S_t)$ differs across periods if S changes, the agent becomes progressively more easily satisfied.

Log-linearizing (and only here taking S approximately constant):

$$\Delta \ln c_{t+1} \approx \frac{1}{\sigma(S)} \left[\ln \beta(S) + r_{t+1}\right]$$

Under the conditions of Proposition 2 (patience effect dominant, satiation condition holding on the relevant domain), the effect on the *level* of consumption is downward: higher S means lower c and higher saving. The qualifier matters; the generosity channel pulls resources toward giving, and it is the dominance condition, verified under the baseline calibration, that signs the total.

B Extended Proofs

B.1 Proof of Proposition 1: The threshold policy for corruption

By the honesty Kuhn–Tucker conditions of Appendix A, the Lagrangian is linear in θ_t, so the optimal policy compares the spiritual cost of corrosion with the rents on offer. Define the cost–benefit ratio

$$h(S) \equiv \frac{\tilde{\mu}^S(S)\,\rho}{\tilde{\lambda}(S)\,\Pi}.$$

The policy is $\theta^* = \bar{\theta}(1 - S)$ when $h(S) < 1$ and $\theta^* = 0$ when $h(S) > 1$, with indifference at $h = 1$. By the proposition's assumption h is nondecreasing in S, so the set $\{S : h(S) \geq 1\}$ is an upper interval; let S^θ be its infimum. Below S^θ the ceiling binds and $\partial\theta^*/\partial S = -\bar{\theta} < 0$; at S^θ the policy drops from $\bar{\theta}(1 - S^\theta)$ to zero; above it the policy is identically zero with derivative zero. The policy is therefore weakly decreasing everywhere and strictly decreasing wherever it is positive, which is the claim.

Two remarks keep this honest. First, the monotonicity of h is an assumption, not a theorem. The economics behind it is the patience channel: $\tilde{\mu}^S$ carries the factor $\beta(S_{t+1})$ and the channel terms of equation (22) strengthen with S, so the numerator rises. But the denominator moves too ($\tilde{\lambda} = c^{-\sigma(S)}$ rises as consumption falls), so the ratio's direction is a joint restriction, verified under the baseline calibration rather than derived from primitives. Second, the strict global inequality $\partial\theta^*/\partial S < 0$ of earlier drafts was too strong: at the zero-corruption corner the derivative is zero, and the honest statement is the weak one with strictness where $\theta^* > 0$. □

B.2 Proof of Proposition 3: g/c is increasing in S

Let $R(S) \equiv \psi(S)^{1/\sigma(S)}$. Take logs:

$$\ln R = \frac{\ln(\psi_0 S^\eta)}{\sigma(S)} = \frac{\ln\psi_0 + \eta \ln S}{\sigma(S)}$$

Differentiating with the quotient rule, using $\sigma'(S) = \bar{\sigma} - \underline{\sigma} \equiv \Delta\sigma$:

$$\frac{R'}{R} = \frac{1}{\sigma}\left[\frac{\eta}{S} - \frac{\ln\psi}{\sigma}\,\Delta\sigma\right]$$

The first term $\eta/(S\sigma)$ is always positive. The second involves $\ln\psi = \ln\psi_0 + \eta \ln S$. Whenever $\psi \leq 1$ we have $\ln\psi \leq 0$, so $-\ln\psi\,\Delta\sigma/\sigma$ is also non-negative and $R' > 0$. Under the baseline calibration ($\psi_0 = 0.05$, $\eta = 1$) this holds *globally*: $\psi(S) = 0.05\,S \leq 1$ for all $S \in [0, 1]$, so g/c is strictly increasing in S on the entire admissible domain. (More generally, any $\psi_0 \leq 1$ gives $\psi \leq 1$ throughout and a global result; only for $\psi_0 > 1$ at high S does the second term turn negative, and it is then of order $\ln S/\sigma^2$, dominated by $\eta/(S\sigma)$.) □

B.3 Proof of Proposition 5: The two thresholds of the forgiveness game

Dominance requires *two* payoff comparisons, one against each strategy of the opponent, and a single threshold cannot carry both. Define the gain to forgiving against a forgiver and against

a revenger:

$$\Delta_F(S) \equiv \pi^{FF}(S) - \pi^{RF}(S) = \underbrace{\underline{\beta}(S)V^{\text{peace}} + \psi(S)(\Omega+\Gamma)}_{\equiv \Phi(S)} - R_0,$$
$$\Delta_R(S) \equiv \pi^{FR}(S) - \pi^{RR} = -L + \beta(S)V^{\text{recovery}} - \pi^{RR}.$$

Both are strictly increasing and continuous: $\Phi'(S) = (\bar{\beta} - \underline{\beta})V^{\text{peace}} + \psi_0\eta S^{\eta-1}(\Omega+\Gamma) > 0$ and $\Delta_R'(S) = (\bar{\beta} - \underline{\beta})V^{\text{recovery}} > 0$. If $\Delta_F(0) < 0 < \Delta_F(1)$, the intermediate value theorem gives a unique $S^F \in (0,1)$ with $\Phi(S^F) = R_0$; likewise a unique S^R solves $\Delta_R = 0$ when Δ_R changes sign on $(0,1)$, and we set S^R below zero (above one) when Δ_R is everywhere positive (negative).

The equilibrium map then reads off the definitions. (F,F) is a Nash equilibrium if and only if $\Delta_F(S) \geq 0$, that is, $S \geq S^F$. (R,R) is a Nash equilibrium if and only if $\Delta_R(S) \leq 0$, that is, $S \leq S^R$. Forgiveness is strictly dominant when both gains are positive, $S > \max\{S^F, S^R\}$, and revenge when both are negative, $S < \min\{S^F, S^R\}$. Between the thresholds, the ordering decides the regime: $S^F < S^R$ yields the coordination case (both symmetric profiles are equilibria and expectations select between them), while $S^R < S^F$ yields hawk–dove (neither symmetric profile survives; the pure equilibria are the asymmetric (F,R) and (R,F), plus a mixed one).

Under the baseline calibration ($\underline{\beta} = 0.92$, $\bar{\beta} = 0.98$, $V^{\text{peace}} = 6$, $R_0 = 7$, $\psi_0(\Omega+\Gamma) = 5$, $\eta = 1$, $L = 3$, $V^{\text{recovery}} = 5$, $\pi^{RR} = -2$): $\Phi(S) = 5.52 + 5.36\,S$, so $S^F = 1.48/5.36 \approx 0.28$, above the low-soul equilibrium $\bar{S}_L \approx 0.20$. And $\Delta_R(S) = 3.6 + 0.3\,S > 0$ for every S, so S^R sits below zero: recovery after victimization beats mutual destruction at any soul state, mutual revenge is never an equilibrium, and revenge is never strictly dominant. The operative regime below S^F is therefore hawk–dove. What fails in the low-soul society is *mutual* forgiveness: every equilibrium there has one group taking revenge while the other's forbearance is exploited. Had the victimization loss been deep enough ($L > 6.6$ at these parameters), S^R would enter the unit interval and mutual revenge would become self-fulfilling in the intermediate range, the coordination trap. Either way the proposition's regime map applies; only the ordering of the thresholds is a calibration fact. □

Remark. Partial spiritual transformation is dangerous, and the hawk–dove regime makes the warning formal: below S^F the equilibria *are* the exploitation profiles, so if only one group crosses the threshold in spirit while the other remains below, the forgiving group gets preyed upon in equilibrium, not by accident. This is why community-wide spiritual renewal is necessary; individual virtue without collective transformation leads to victimization.

B.4 Proof of Proposition 7: Multiple equilibria

With social interaction, the steady-state condition is:

$$(\delta^S - \xi)\bar{S}^* = \phi^S\big(i^{S*}(\bar{S}^*)\big) + \bar{G} - \rho\,\theta^*(\bar{S}^*)$$

Define $L(\bar{S}) = (\delta^S - \xi)\bar{S}$ (linear, positive slope under the condition $\delta^S - \xi > 0$) and the *replenishment curve*

$$R(\bar{S}) = \phi^S\big(i^{S*}(\bar{S})\big) + \bar{G} - \rho\,\theta^*(\bar{S}).$$

The shape assumption, stated as such. The proposition assumes R is continuously differentiable, nondecreasing, and S-shaped relative to the depreciation slope: there exist $0 < s_1 < s_2 < 1$ with

$$R'(\bar{S}) < \delta^S - \xi \ \text{ on } [0, s_1), \qquad R'(\bar{S}) > \delta^S - \xi \ \text{ on } (s_1, s_2), \qquad R'(\bar{S}) < \delta^S - \xi \ \text{ on } (s_2, 1],$$

together with the endpoint conditions $R(0) > 0$ and $R(1) < (\delta^S - \xi)$. We are explicit that this is an *assumption* about the composite object R, because the primitives alone do not deliver it: concavity of ϕ^S restricts the production function, not the optimal policy $i^{S*}(\bar{S})$ composed with it, and a merely convex or merely concave R crosses a line at most twice. The economic case for the shape is the investment complementarity, and it runs regime by regime: (i) **flat** at low $\bar{S}$, where the social environment is hostile, $\tilde{\mu}^S$ is small (the honesty drag $-\tilde{\nu}\bar{\theta}$ is deepest here), and $i^{S*} \approx 0$; (ii) **steep** at intermediate $\bar{S}$, where higher $\bar{S}$ raises $\tilde{\mu}^S$ through the channels and the social multiplier, which raises i^{S*}, which reinforces $\bar{S}$, while the receding corrosion term $-\rho\theta^*(\bar{S})$ adds slope as cheating recedes; (iii) **flattening** at high $\bar{S}$, where the concavity of ϕ^S finally dominates and $\theta^* \to 0$ leaves nothing more to recover. The calibrated policy function satisfies the assumption (the logistic investment response of Figure 2); a derivation of the S-shape from primitives would strengthen the proposition and we flag it as open.

Three crossings. Define $H(\bar{S}) \equiv R(\bar{S}) - L(\bar{S})$. The endpoint conditions give $H(0) > 0$ and $H(1) < 0$, and the slope pattern makes H strictly decreasing on $[0, s_1)$, strictly increasing on (s_1, s_2), and strictly decreasing on $(s_2, 1]$. A function with this decreasing–increasing–decreasing profile and these endpoint signs has exactly one zero in each monotone segment where it changes sign, and the signs force exactly three zeros: $\bar{S}_L < \bar{S}^* < \bar{S}_H$, with $\bar{S}_H$ interior (below the boundary $\bar{S} = 1$) precisely because the final flattening regime pulls R back under L before the boundary.

Stability against the discrete dynamics. The mean soul state evolves by $\bar{S}_{t+1} = F(\bar{S}_t)$ with

$$F(\bar{S}) = \bar{S} + R(\bar{S}) - (\delta^S - \xi)\bar{S}, \qquad F'(\bar{S}) = 1 + R'(\bar{S}) - (\delta^S - \xi).$$

Local stability of a fixed point requires $|F'| < 1$, that is,

$$(\delta^S - \xi) - 2 \ < \ R'(\bar{S}) \ < \ \delta^S - \xi.$$

The upper inequality is the familiar slope comparison $L' > R'$. The lower inequality is the one a pure slope comparison forgets (it rules out overshooting oscillations), and it is where the monotonicity restriction on R earns its place: since $R' \geq 0$ and $\delta^S - \xi < 2$ under any sensible calibration, the lower bound holds automatically, and for monotone replenishment the slope comparison is the complete stability condition. It holds at $\bar{S}_L$ and $\bar{S}_H$ (both in slow-R regimes) and fails at $\bar{S}^*$ (in the steep regime), so the outer steady states are locally stable and the middle one is the unstable tipping point. □